Universidad de Chile
Escuela de Derecho
Contratos Parte Especial
Prof. Francisco González Hoch

Materiales Optativos I

"Evolución de las formas contractuales y fundamentos de la obligatoriedad de los contratos"

1. ATIYAH, Patrick S. *Una Introducción al Derecho de Contratos*. [trad. María Agnes Salah] Clarendon Law Series, Clarendon Press – Oxford (1995) pp. 43

2. ATIYAH, Patrick S. *Forma y substancia en el derecho de contratos.* [trad. Rodrigo Soto Silva] Clarendon Press – Oxford (1984) pp. 93-120

3. MALINOWSKY, Bronislaw. *Crimen y Costumbre en la Sociedad Salvaje*. Barcelona: Editorial Ariel (1956 [1926]) pp. 27-73

4. GEHLEN, Arnold. *Antropología Filosófica*. Barcelona: Paidós (1993) pp. 36-40; 87-95

5. HORWITZ, Morton J. *The Historical Foundations of Modern Contract Law*. Harvard Law Review, Vol. 87, No. 5 (Mar., 1974) pp. 917-956

6. PIZARRO WILSON, Carlos. *Notas críticas sobre el fundamento de la fuerza obligatoria del contrato. Fuentes e interpretación del artículo 1545 del Código civil chileno*. Revista Chilena de Derecho, Pontificia Universidad Católica de Chile, vol. 31, no. 2 (2004) pp. 225-237

7. CRASWELL, Richard. *Contract Law, Default Rules, and the Philosophy of Promising.* Michigan Law Review, Vol. 88, No. 3 (Dec., 1989) pp. 489-529

8. SHIFFRIN, Seana Valentine. *The Divergence of Contract and Promise*. Harvard Law Review, Vol. 120, No. 3 (Jan., 2007) pp. 708-753

9. LIGHTSEY, Wallace K. *A Critique of the Promise Model of Contract*. William and Mary Law Review, Vol. 26, Issue 1 (1984) pp. 45-73

10. HERMANLIN, Benjamin; KATZ, Avery; CRASWELL, Richard. *Chapter on the Law & Economics of Contracts.* Forthcoming in The Handbook of Law & Economics (final draft: June 5, 2006) pp. 143

11. POSNER, Eric. *Economic Analysis of Contract Law after Three Decades: Success or Failure?* The Yale Law Journal, Vol. 112, No. 4 (Jan., 2003) pp. 829-880

Una introducción al

DERECHO

DE.CONTRATOS

P. S. ATIYAH

EX-PROFESOR DE DERECHO INGLES
DEL ST JOHN'S COLLEGE EN LA UNIVERSIDAD DE OXFORD

(quinta edición. 1995)
Clarendon Law Series
Clarendon Press - Oxford

La traducción ha sido preparada por la alumna María Agnes Salah y revisada por Andrés Jana L.

EL DESARROLLO DEL DERECHO DE CONTRATOS MODERNO.

1. LOS CONTRATOS Y EL DERECHO DE OBLIGACIONES.

El derecho de contratos forma parte del derecho de obligaciones , es decir, tiene relación con las obligaciones que las personas contraen para con otros como resultado de las relaciones y transacciones en las que se ven envueltos. Más generalmente, el derecho de contratos es parte del derecho privado, en el sentido que las obligaciones de carácter públicas, - como las obligaciones constitucionales o políticas, no son tratadas por el derecho, o concebidas por los abogados, como parte del derecho de obligaciones. Es cierto que los entes públicos pueden celebrar contratos comunes y de este modo - someterse al derecho privado, pero los deberes más amplios de estos entes no caen dentro del campo del derecho de obligaciones como es entendido comúnmente. Así también, el derecho penal no es concebido por los abogados como parte del derecho de obligaciones. Desde luego que el derecho penal impone deberes en los ciudadanos, y estos deberes son, en un sentido, obligaciones legales. Pero estos deberes no son debidos a nadie en particular, y su ejecución normalmente descansa en la policía y en otros entes públicos. Por el contrario, el derecho de obligaciones dice relación primeramente con obligaciones debidas por algunos miembros de la sociedad para con otros, y estas obligaciones pueden ser hechas cumplir exclusivamente por aquellas personas a quienes son debidas. Una persona que ha sido víctima de un crimen puede reclamar a la policía, quien investigará, y si lo considera apropiado podrá enjuiciar al ofensor. Pero una persona que desea reclamar por el incumplimiento de una obligación que le es debida, tal como el incumplimiento de un contrato, debe hacer efectivos sus derechos en los tribunales, sin la asistencia de autoridad pública alguna.

Las obligaciones emanan de una variedad de fuentes, y pueden ser clasificadas de diversas formas. Pueden, por ejemplo, ser clasificadas de acuerdo a las relaciones sociales de las cuales emanan. De este modo, uno puede distinguir entre obligaciones debidas por una persona a los miembros de su familia, obligaciones entre vecinos, obligaciones que emanan de una relación laboral, etc.. Sin embargo, ha sido tradicional en el derecho tratar como distinción básica la que diferencia entre las obligaciones que son adquiridas por el individuo por su propia voluntad y las obligaciones que son impuestas al ciudadano desde

fuera. Hablando en terminos más generales, el derecho de contratos es aquella parte del derecho que dice relacion con las obligaciones que el individuo adquiere por su propia voluntad. Otras áreas importantes del derecho de obligaciones son el derecho de daños extracontractuales (-que en términos generales dice relación con los delitos civiles y perjuicios extracontractuales) y las acciones restitutorias (-que en términos generales dice relación con la obligación restitutoria de dinero o de pagar por beneficios recibidos cuando no ha habido contrato que justifique el pago de estos beneficios o dinero). Como será visto más adelante, estas distinciones no son del todo absolutas, y uno de los fenómenos más notables de los tiempos modernos ha sido que los límites entre el derecho de contratos y otras partes del derecho de obligaciones se han hecho cada vez más borrosos. En particular, como se discutirá más adelante con detalle, se sugerirá que muchas de las obligaciones reconocidas por el derecho de contratos no pueden ser concebidas en realidad como adquiridas por voluntad propia.

Por el momento estos matices pueden ser dejados de lado, y podemos concentrarnos en el hecho indubitado de que el derecho de contratos permite a las personas imponerse obligaciones sobre sí mismos. Tal orden jurídico presupone naturalmente una sociedad y un sistema legal en el cual las personas tienen el derecho de elegir qué obligaciones desean asumir. En sociedades muy primitivas, el rol del contrato ha sido generalmente muy pequeño, porque las obligaciones son normalmente concebidas como emanaciones de la costumbre y del orden social antes que de una elección libre. Igualmente en las sociedades colectivas modernas, donde el Estado es todopoderoso y los derechos individuales de libre elección son menos respetados, el rol del derecho de contratos puede ser al menos en la práctica, menos significativo. Pero en las Democracias Occidentales, donde tradicionalmente son respetados derechos más amplios en lo referente a la libre elección, el derecho de contratos ha jugado un rol principal. En el desarrollo del *common law* inglés, las ideas contractuales adquirieron mayor importancia a partir del siglo dieciséis a medida que el mayor individualismo y libertad de la Inglaterra de post-Reforma fueron comenzando a establecerse. No cabe duda que el derecho ha sido influenciado por una considerable cantidad de factores, pero probablemente no es exagerado decir que dos de estos factores han sido lejos de mayor importancia que cualquier otro. Estos son el factor moral y el factor económico o comercial.

A pesar de que los abogados y teóricos ingleses están tradicionalmente habituados a insistir en el derecho y la moral son distintos, no es menos cierto que el derecho refleja en una medida considerable los estándares morales e

ideales de la comunidad en la cual opera. Es por esto que no es sorprendente encontrar que tras una gran parte del derecho de contratos, descansa el simple principio moral de que una persona debe cumplir sus promesas y respetar sus acuerdos. Esto no quiere decir que el antiguo derecho inglés convirtió este principio moral en una norma legal, porque no fue, de hecho, sino hasta finales del siglo dieciséis que adquirimos algo semejante a un derecho general de los contratos, y cuando esto ocurrió fue principalmente bajo el ímpetu del factor comercial o económico. Más aún, como se verá más adelante, todavía existen dudas y controversias acerca de si el derecho realmente considera el incumplimiento de contrato como algo ilícito, y algunas partes del derecho parecen tolerar la visión de que efectivamente no hay nada malo en incumplir un contrato mientras los daños causados por cualquier pérdida sean pagados. Y una vez más, se necesitaría hacer todo tipo de calificaciones a la idea de que la obligación moral de cumplir una promesa es una idea tan simple como parece. Aún, por el momento, es suficiente notar que al menos una fuerte corriente del derecho contractual deriva de la idea de que una persona debe cumplir su palabra, y que las promesas imponen obligaciones morales.

Con el desarrollo económico y social de las sociedades modernas, la necesidad de un derecho de contratos se hace aún más urgente por al menos dos razones. En primer lugar, la división del trabajo, que es un rasgo fundamental de las sociedades modernas, crea una creciente demanda para la transferencia de propiedad de algunos miembros de la comunidad a otros y para la prestación de servicios de algunos miembros de la comunidad para otros. El mecanismo legal por el cual estas transferencias de propiedad y prestación de servicios son llevadas a cabo es, en términos generales, el derecho de contratos. El derecho contractual es entonces, en gran parte, el derecho de los intercambios, el derecho que regula los métodos por los cuales los individuos intercambian bienes y servicios, usualmente a cambio de dinero.

2. CONTRATOS E INTERCAMBIO ECONÓMICO:

Existen, por supuesto, muchas formas de intercambio. El derecho contractual está relacionado ampliamente con el intercambio económico que tiene lugar en el mercado, con compras y ventas, contrataciones y arrendamientos, empleos y servicios, préstamos de dinero, etc.. Existen también otras formas de intercambio, como las que tienen lugar en el contexto familiar ("Tú haces el

lavado y yo haré las compras" o "Yo pagaré la hipoteca y tú pagarás las cuotas pendientes del auto"). El derecho contractual tiene poca injerencia en aquellos intercambios que no son de mercado, aunque ocasionalmente puede ser invocado cuando, por ejemplo, una pareja no casada compra una casa en conjunto y posteriormente se separa. Cuando se trata de parejas casadas, la disputa litigiosa es, por razones obvias, improbable a menos que se separen o se divorcien, o que alguno de ellos muera y luego haya disputas respecto de la titularidad de la propiedad de él o de ella. Esas disputas no son usualmente resueltas invocando al derecho contractual, sino por medio de diversas normas legales, por cierto frecuentemente bajo un régimen que en gran medida descansa en la discrecionalidad judicial antes que en normas. Pero incluso en casos de este tipo muchas de las normas e ideas que están insertas en el derecho de contratos serán utilizadas a menudo - por ejemplo, en el caso de jueces que haciendo uso de su discrecionalidad para repartir la propiedad matrimonial en casos de divorcio, pueden ser influenciados por ideas de "intercambio justo" (¿Qué aportó cada cual?) aunque por supuesto, las necesidades familiares y otras consideraciones también pueden ser relevantes. En los sistemas económicos occidentales, el intercambio económico, con el que el derecho contractual está particularmente relacionado, es considerado generalmente como un importante instrumento de eficiencia económica en dos formas principales. En primer lugar, el intercambio libre y voluntario, es generalmente un simple, pero críticamente importante, método de incremento de la satisfacción del consumidor, e incluso de incremento de la riqueza de la comunidad, si es que la riqueza es definida (como frecuentemente lo es) en términos de satisfacción del consumidor. Cuando dos partes acuerdan libremente un contrato en que está envuelto, digamos por ejemplo, un simple intercambio de dinero por bienes, el vendedor realiza el intercambio porque piensa que será más rico con el dinero que con los bienes, y el comprador lo realiza porque prefiere tener los bienes más que el dinero. Ambas partes, entonces, emergen del intercambio "más ricos" (en un sentido) de lo que eran antes, y, como la riqueza de la sociedad está construida a partir de la riqueza del total de sus miembros, incluso un simple intercambio de este tipo puede aumentar la "riqueza" social. De hecho, excepto para alguien que lleve una vida como la de Robinson Crusoe, los más simples intercambios son una necesidad absoluta para todos. Pocas personas en una sociedad moderna podrían sobrevivir sin intercambiar su trabajo por dinero, y su dinero por bienes y servicios. El más simple de los objetos que compramos en un mercado frecuentemente es, en sí mismo, el producto de un sin fin de previos intercambios.

Comprar un lápiz en una tienda, y considerar quien cultivó el bosque, como fue talada la madera, llevada a un puerto, transportada en un buque (construido a su vez en base a innumerables intercambios), enviada a un astillero, cortada nuevamente, manufacturada para convertirla en un lápiz, distribuida a los comerciantes, y finalmente ofrecida para la venta a los consumidores.

Por supuesto, muchos intercambios son mucho más complejos que simples ventas : un contrato de empleo, por ejemplo, que continúa en vigencia a lo largo de un período de tiempo, es mucho más complejo que una simple venta de bienes; y obviamente, las grandes transacciones relacionadas con (por ejemplo), la construcción de grandes proyectos de ingeniería, o la venta de cargamentos, o el alquiler de naves para que transporten dichos cargamentos, significarán a menudo acuerdos muy complicados. En el mundo moderno, el número de intercambios diarios es enorme y su valor total astronómico. Por la misma razón, la ganancia total en la riqueza de todos los habitantes del mundo que deriva del libre intercambio económico, es casi inimaginable. Este es, en efecto, el motivo por el cual es posible la existencia del comercio internacional. Semejante comercio es obviamente muy costoso. Los barcos deben ser construidos, equipados y dotados de personal a un gran costo. Pero las ganancias que se producen al acarrear mercaderías a lo largo de los mares y permitir que sean vendidas bajo un sistema de libre intercambio, son tan enormes, que alcanza para pagar todos estos costos e incluso queda algo de sobra. Un momento de reflexión es, por lo tanto, suficiente para demostrar cuan importante es para cualquier sociedad el intercambio libre y voluntario. Interferir en él es algunas veces necesario; pero para hacerlo debemos estar seguros de que *es* necesario, y además debemos considerar los costos de hacerlo.

Existe una segunda razón de por qué el intercambio libre y voluntario es un instrumento principal de eficiencia económica. Es el libre intercambio quien determina en gran medida cómo los recursos de la sociedad deben ser asignados entre diferentes usos posibles. En una sociedad de libre empresa no es el Estado ni el gobierno quien decide cuántos autos deberían producirse (y cuántos de qué modelo en particular), cuanto dinero debería ser destinado a la industria de entretenimiento, o si los supermercados o los almacenes deben abrir en un nuevo lugar, etc.. Estas son materias que están para ser determinadas por el mercado, operando a traves del libre intercambio, y por ende a través de contratos. Es a los consumidores a quienes corresponde la ultima elección. Son los consumidores quienes determinan, eligiendo comprar una cosa en vez de otra, en un lugar en vez de otro, bajo determinados términos en vez de otros, qué recursos deberían

destinarse a la producción y distribución de qué bienes y servicios. Si los consumidores son libres de realizar la elección que ellos deseen - obviamente dentro de los límites de sus propios recursos - y los abastecedores son libres de responder a la demanda de los consumidores, entonces el resultado, en teoría al menos, es que el mercado podrá a largo plazo ofrecer aquello que los consumidores en su conjunto han mostrado que desean, en las proporciones en que lo desean y a los precios a que están dispuestos a pagar. Es ésta una razón fundamental de porqué la libertad de elección en el intercambio - libertad contractual - es estrechamente asociada con el alza (y algunas veces con la caída) de la creencia en el libre mercado.

Pero el solo intercambio no es suficiente. Únicamente las transacciones más simples pueden consumarse por un intercambio simultáneo, como el que tiene lugar en la caja de un supermercado, donde el comprador descarga su carro y entrega su dinero. La mayoría de los intercambios, sea cual sea su complejidad, no pueden ser realizados en forma simultánea. Una o ambas partes deberán ejecutar en el futuro lo acordado, lo que significa que una parte debe *confiar* en la ejecución de la otra, debe tener confianza en que cumplirá. Por lo tanto, la mayoría de los contratos requieren cierto grado de cooperación y confianza, y, una sociedad moderna no puede existir sin un alto grado de actividad cooperativa. Gran parte de esta cooperación es regulada por leyes de derecho público más que por el derecho de contratos - por ejemplo, el mantenimiento de la ley y el orden es garantizado por el establecimiento de fuerzas policiales, cortes criminales, cárceles, etc.. Pero una gran cantidad de actividad cooperativa también es garantizada por el intercambio voluntario, sin el cual sería simplemente imposible agrupar el capital y la fuerza laboral necesaria para proyectos industriales de gran escala - a menos que, por supuesto, fuera realizado mediante la compulsión del poder del Estado. Pero el intercambio voluntario a esta escala también nos conduce a una necesidad de confianza. En el proceso de transferencia de la propiedad y de prestación de servicios, las personas por necesidad deberán confiar fuertemente en las promesas y los acuerdos. La mayoría de los acuerdos para la transferencia de mercaderías, bienes raíces o servicios no pueden ser realizados instantánea y simultáneamente. Muchas veces estos acuerdos deben planearse por anticipado. Casi siempre una parte debe cumplir antes que la otra, y a menudo los acuerdos deben ser cuidadosamente realizados por anticipado, antes que cualquier ejecución pudiese tener lugar. En una sociedad razonablemente ordenada y estable, hacer estos acuerdos anticipados genera expectativas acerca de la

conducta futura de las otras partes. En las sociedades modernas todos tenemos expectativas razonables acerca de como otros se comportarán, y todos dependemos, o confiamos, en que los otros se comportarán en el futuro como han dicho que lo harán.

En este sentido, el derecho contractual es un instrumento para garantizar la cooperación en el comportamiento humano y particularmente en el intercambio. "A" consiente en rascar la espalda de "B" si "B" accede a hacer lo mismo cuando "A" se lo pida. Pero a menos que "A" esté razonablemente confiado en que "B" le corresponderá cuando llegue el momento, "A" se mostrará reacio en ayudar a "B" aquí y ahora. Una forma de hacer que "A" tenga más confianza en que "B" le corresponderá, es hacer que "B" prometa que así lo hará. Pero incluso eso no es todavia suficiente por sí mismo, ya que las promesas pueden romperse. Por supuesto que existen muchas sanciones para los quebrantadores de promesas, y el derecho no es necesario para todas ellas. La sanción más simple consiste en no tratar con esa persona nuevamente. Así incluso en áreas donde no existe un derecho real y efectivo que permita hacer cumplir forzadamente las promesas y los contratos, existen poderosas razones por las cuales las promesas y los contratos tienden a ser cumplidos. Por ejemplo, considere cuan extraño es ver que gobiernos que piden enormes sumas de dinero prestado en los mercados financieros mundiales incumplan sus obligaciones, no obstante que no existe una forma real de obligarlos a cumplir sus contratos. Ellos pagan porque saben que si no lo hacen, tal vez no podrán nunca más volver a solicitar un préstamo. Sin embargo, algunas veces las sanciones de este tipo, aunque son mejor que nada, no son satisfactorias, hay ocasiones en que la cooperación no existira porque las partes saben que no hay sanciones en caso de incumplimiento. Algunas veces esto es en realidad algo positivo, cuando por ejemplo, estamos tratando con contratos no deseables, como el chantaje y los acuerdos de corrupcion. Porque los acuerdos de este tipo usualmente no son exigibles legalmente, las partes que deseen realizar semejantes acuerdos tendrán poca confianza en que la otra parte cumplirá como fue acordado. Esta falta de confianza hace aún más difícil la realización de tales acuerdos, y es a menudo necesario para el chantajista o el sobornador tener que arreglar alguna forma de intercambio simultáneo, ya que no está dispuesto a cumplir primero, dejando libre a la otra parte para incumplir más tarde. Pero el intercambio simultáneo no es siempre practicable, de modo que estos intercambios no deseables son frecuentemente difíciles de llevar a cabo - y esto es por supuesto, algo positivo.

Sin embargo, la mayoría de los intercambios libres y voluntarios son socialmente deseables : queremos facilitarlos no desalentarlos. Y en la ausencia de sanciones legales para forzar su cumplimiento, las partes se verán reducidas a procedimientos menos satisfactorios, como en el caso de los intercambios no deseables. De este modo, obtenemos el tradicional intercambio de capa y espada de espías capturados, donde los dos espías caminan a lo largo de un puente o frontera mientras los guardias aguardan al final de cada lado. Este es un intercambio deseado, con todo, es uno de esos intercambios que no puede ser respaldado fácilmente por una sanción legal, así que las partes están forzadas a hacer el intercambio simultáneamente. Esto difícilmente sería deseable o practicable para la gran cantidad de intercambios económicos ordinarios, y es precisamente por esta razón que necesitamos auténticas sanciones legales. Por consiguiente, la mayoría de los sistemas legales han establecido reglas que impondrán sanciones en aquellos que violen sus contratos. La cooperación entonces se transforma en algo más sencillo, y los intercambios se facilitan.

En las sociedades modernas sofisticadas esta cooperación ha conducido a un masivo y elaborado sistema de crédito - y "crédito" es simplemente otra palabra para "confianza" o "seguridad ". El crédito es, en efecto, seguridad o confianza en la parte aún no realizada de un acuerdo o intercambio. En el tipo de caso más simple, cuando una persona provee bienes o servicios a crédito a un consumidor, confía en que el consumidor pagará, y en el intertanto, permite al consumidor tener los bienes. Generalmente el consumidor al final pagará, pero si no lo hace, se necesita alguna sanción, y la sanción es proporcionada por el derecho de contratos. De este modo, el derecho de contratos es quien finalmente otorga el respaldo necesario para sostener a toda la institución del crédito. Una rápida reflexión es suficiente para demostrar hasta qué punto esto es cierto, no sólo en materias comerciales, sino en todas las instancias de la vida. Una cuenta bancaria de un consumidor, el derecho de habitar su casa si es arrendada o hipotecada, su empleo, su seguro, sus acciones y muchas otras materias de vital importancia para él, todas depende, para su valor de el hecho que, en última instancia, el derecho de contratos le permitirá ejercer sus derechos. En la notable frase de Roscoe Pound : "La riqueza, en una era comercial, está hecha mayormente en base a promesas." (1) Esta es la razón de porqué el desarrollo del derecho de contratos, tanto en Inglaterra como en cualquier otro lugar, ha sido tan ampliamente asociado con el desarrollo del comercio.

3. EL DERECHO CONTRACTUAL CLÁSICO.

Aunque gran parte del derecho de contratos inglés tiene raíces que se remontan a la Edad Media, la mayoría de los principios generales del derecho moderno fueron desarrollados y elaborados en los siglos dieciocho y diecinueve. Estos principios, y quizás aún más, la aproximación general de las cortes a cuestiones contractuales, no sería impropio atribuirlo a la tradicional, o clásica, teoría del derecho de contratos. El derecho de contratos ha experimentado algunos cambios fundamentales en el siglo veinte, pero es absolutamente imposible comprender el derecho moderno sin algun conocimiento y apreciación del trasfondo y los orígenes del derecho clásico. (2)

Los siglos dieciocho y diecinueve representaron el apogeo de las teorías del derecho natural y la filosofía del *laissez-faire*, y muchos de los jueces, quienes fueron largamente responsables de la creación del derecho de contratos durante este período, fueron, como la mayoría de los hombres instruidos de la época, considerablemente influenciados por los pensamientos del momento. Para los jueces del siglo dieciocho las teorías del derecho natural significaban que los hombres poseían un derecho inalienable a la propiedad, y por lo tanto, también para realizar sus propios acuerdos, para comprar o vender o para distribuir esa propiedad y por consiguiente, para celebrar contratos por sí mismos. Al mismo tiempo el derecho estaba todavía influenciado por el paternalismo, que fue una de las características del siglo XVIII. Así, parte de la potencial severidad de un regimen de libertad contractual estricto, fue mitigado durante este período por doctrinas y reglas paternalistas que permitieron a los jueces proteger a aquellos menos capaces de valerse por si mismos en el libre mercado.

Durante el siglo diecinueve las ideas paternalistas languidecieron en la medida que la filosofía del *laissez-faire* tomó fuerza. La mayoría de la gente instruida, incluidos los jueces, entendió el *laissez-faire* con el significado que el derecho debería interferir con las personas lo menos posible. Para los jueces la función del derecho privado comenzó a ser vista mayoritariamente como una negativa. Su objetivo principal fue permitirle a las personas "cumplir sus deseos ", o, en un lenguaje más prosaico, dejarlos seguir adelante con sus negocios, conducir sus asuntos comerciales como lo estimaran mejor, dejarlos guiar sus vidas desembarazados de la interferencia del gobierno, etc...

De una manera general, esto significó que el derecho de contratos fue diseñado para ser el sostén de la ejecución de los convenios privados que las partes contratantes habían acordado. En general, el derecho no se ocupaba de la

equidad o la justicia del resultado y las ideas paternalistas comenzaron a ser consideradas como anticuadas . Los jueces no estaban siquiera muy preocupados de la posibilidad de que el contrato no fuera en el interés público. Así, la función del derecho contractual era simplemente asistir a una de las partes contratantes cuando la otra rompía las reglas del juego y dejaba de cumplir sus obligaciones contractuales. El juez era sólo una especie de árbitro cuyo trabajo era responder al llamado cuando algo andaba mal. Estas ideas significaron alentar casi ilimitadamente la libertad para contratar, y de este modo los lemas "libertad contractual" e "inviolabilidad de los contratos" se convirtieron en las bases sobre las cuales fue construido todo el derecho de contratos. Sería, sin embargo, erróneo concluir que los jueces de este período no estaban interesados en la justicia : ellos pensaban que era *justo* hacer cumplir los deberes contractuales en forma estricta de acuerdo a la letra. Ellos llegaron a pensar que ese paternalismo era proteccionista para los individuos, quienes eran suficientemente fuertes para sostenerse por ellos mismos en el mercado. Sería igualmente erróneo pensar que los jueces eran indiferentes al interés público. Ellos simplemente pensaban que, en casi todos los casos, *era* de interés público hacer cumplir los contratos privados. En efecto, era ampliamente pensado que esto estaba demostrado por principios económicos fundamentales(3).

Es conveniente evitar una simplificación extrema. No todos los jueces en la Inglaterra del siglo diecinueve eran entusiastas adherentes de la libertad contractual. Pero a todo nivel, aproximadamente después de 1830, la ideología del *laissez-faire* tuvo una significativa influencia en el desarrollo del derecho contractual. En particular, muchas "doctrinas equitativas"* invocables en la Court of Chancery** , y diseñadas para proteger a aquellos que celebraran negocios tontos y descuidados, comenzaron a ser cercenadas por los jueces. El paternalismo de los jueces del siglo dieciocho fue ampliamente repudiado por sus sucesores del siglo diecinueve. Este rechazo del paternalismo era en realidad parte de un movimiento de reforma que fue un aliado cercano del movimiento

---

* Las *equitable doctrines* del Derecho Anglosajón son un conjunto de doctrinas que se basan en normas y principios de equidad, esto es conforme a los principios de justicia y corrección; en oposición a las estrictas reglas de derecho. La equidad o *equity* está basada en un sistema de reglas y principios originados en el Inglaterra como una alternativa a las severas reglas del *common law* y que se fundamenta en lo que era justo en la situación particular. La protección bajo este sistema se obtiene en una corte de equidad o en una corte que esté dotada de poderes de equidad (*equity powers*). El término "equidad" denota el espíritu y hábito de justicia y corrección en el trato que regula las relaciones entre los hombres. *Black's Law Dictionary*, p. 484, quinta edición. (*n.t.*)

** Corte que tiene jurisdicción y administra justicia y procedimeintos en conformidad a las formas y principios de equidad (*n.t*).

político hacia la democracia. Fueron los reformadores de la década de1830 quienes proclamaron su fe en el individualismo, su creencia en que podía confiarse en la gente para cuidar de sus propios intereses, ya sea en el mercado o en los tribunales.

Como la mayoría de los lemas, eso de la "libertad contractual" raramente recibió, si es que alguna vez lo hizo, el examen profundo que su importancia merecía, e incluso hoy, bajo ningún respecto, es fácil decir qué fue exactamente lo que quisieron expresar los jueces del siglo diecinueve cuando usaban esta frase. Al menos se puede decir que la idea de la libertad contractual abrazó dos conceptos aunque distintos cercanamente conectados, . En primer lugar, enfatizó que los contratos se basaban en el mutuo acuerdo, en tanto que en segundo lugar, acentuó la idea de que la creación de un contrato era el resultado de la libre elección desembarazada de control externo, tales como la interferencia gubernamental o legislativa.

Decir que los contratos se basan en el acuerdo o consentimiento mutuo, es una afirmación que posee apoyo general, y, en efecto, es tomada como axiomática por la mayoría de los abogados. La mayoría de los contratos, *son* creados como resultado del acuerdo de las partes, al menos en cuanto a lo esencial. Sin embargo, son necesarias. algunas calificaciones para la simple afirmación de que los contratos se basan en el acuerdo de voluntades. Primeramente, uno de los rasgos más fundamentales del derecho de contratos es que la prueba del consentimiento es objetivo y no subjetivo. No importa si las partes realmente consintieron en lo más profundo de sus mentes. La pregunta no es si las partes realmente consintieron, sino que si su conducta y lenguaje son tales que conducirían a gente razonable a asumir que la persona consintió. De este modo el fallo del jurado en el clásico caso de *Bardell vs. Pickwick* , aunque discutible, fue correcto ya que el lenguaje del señor Pickwick fue, a lo menos, capaz de ser entendido por una persona razonable como una propuesta de matrimonio, aunque de hecho nada pudo haber estado más lejos de su mente.

Pero más importantes que estos casos de malentendidos tan gruesos, son los innumerables contratos celebrados cada día por los consumidores e incluso por los hombres de negocios, que consisten en formularios impresos y firmados. Está bien establecido en el derecho moderno que los términos impresos de tales contratos son obligatorios para las partes, pese a que pueden no haber sido leídos o entendidos por una o ambas partes (4). Sólo en circunstancias extrañas y excepcionales de fraude u otro grueso error acerca de la naturaleza o efecto fundamental del formulario impreso, puede una parte negar la fuerza de su propia

firma. (5) Ahora, es obvio que las ventajas prácticas de considerar esas firmas como definitivamente obligatorias son muy grandes. Al mismo tiempo es realmente importante tener constantemente presente que estas reglas muchas veces significan que una persona está contractualmente obligada a cumplir cláusulas que simplemente no conocía o no entendió al momento de celebrar el contrato.

Al respecto, el derecho contractual clásico difiere poco del derecho moderno, aunque el enfoque objetivo se ha intensificado con el paso del tiempo. Pero el derecho clásico no se detuvo en este punto. Tan grande fue el énfasis en el acuerdo y en la intención de las partes, que los jueces del siglo diecinueve tendieron a elevar al derecho de contratos a la posición central del derecho de obligaciones en general. Esto condujo a dos desarrollos conexos. Primero, existió una renuencia a imponer obligaciones sobre aquellos que no las habían asumido voluntariamente. El derecho de daños extracontractuales y las acciones restitutorias permanecieron de este modo relativamente poco desarrolladas durante este período. En segundo lugar, cuando eran impuestas obligaciones, hubo una tendencia a asumir que ellas debían ser contractuales. Así por ejemplo, los jueces negaron tener algún poder para "hacer un contrato para las partes". Igualmente, ellos intentaron expresar el gran cúmulo de las reglas contractuales propiamente tales como dependientes de la intención de las partes. Al adoptar esta línea las cortes sintieron que ellos no estaban imponiendo normas legales a las partes, sino que estaban simplemente resolviendo las implicancias sobre lo que las mismas partes habían elegido hacer. Así como John Locke había argumentado que las obligaciones políticas debían su legitimidad al contrato social respecto del cual las personas habían "consentido tácitamente", de la misma forma los jueces argumentaban que las obligaciones privadas dependían principalmente de los contratos privados respecto de los cuales ellos podían encontrar un "consentimiento tácito" o real.

Por supuesto existían ciertas normas legales dominantes respecto de las cuales no podía, en sentido alguno, decirse que dependían de la intención de las partes, tales como las normas relativas a la capacidad contractual de menores de edad, y las normas relativas a los contratos ilegales, pero habían, y hay, otra gran cantidad de normas que las cortes prefirieron considerar como basadas en la intención de las partes. Las siguientes materias, por ejemplo, fueron (y a menudo todavía son) tradicionalmente concebidas como dependientes de la intención de las partes: la cuestión sobre si las partes habían celebrado un contrato; la cuestión sobre si un acuerdo debería tener valor legal como contrato; la

interpretación del lenguaje utilizado por las partes y la cuestión sobre si las afirmaciones hechas por ellas deberían tomarse como contractuales; los efectos legales del contrato, incluso cuando las partes no hubieran declarado expresamente cuales habrían de ser esos efectos; y - quizás lo más extraordinario de todo - la ley de qué país habría de regir el contrato. Más aún, todavía se sigue insistiendo frecuentemente en la idea general de que las obligaciones contractuales derivan del acuerdo o consentimiento mutuo, aún cuando las normas actuales relativas a las acciones disponibles contra el incumplimiento del contrato son rara vez consideradas como descansando sobre la intención contractual. Entonces cada vez que un contrato es incumplido, la acción legal disponible para la otra parte - usualmente una acción por daños - y el tipo de daños que son recuperables, son casi siempre determinados por normas legales y no por la intención de las partes.

Ahora, en cierto sentido fue verdadero (y en menor medida, aún lo es), decir que la mayor parte de las reglas propiamente tales del derecho de contratos se basó en la intención de las partes, porque en la mayoría de los casos estuvo (y sigue estando) abierta a las partes la posibilidad de variar o excluir la aplicación de estas reglas por medio de un acuerdo expreso. Para tomar un simple ejemplo, las obligaciones del vendedor de bienes respecto de la calidad de los bienes vendidos puede variar, o ser excluidas en su conjunto, si así lo acuerdan las partes expresamente, aunque actualmente esto ya no es siempre así. De este modo, para este punto era perfectamente verdadero decir que estas reglas dependían de la intención de las partes. Pero fue claramente un error deducir que porque las partes son libres de variar las consecuencias legales de un contrato, entonces, cuando estas consecuencias no son variadas, son el producto de su acuerdo. Por supuesto, este *puede* ser el caso. Las partes pueden haber examinado las normas legales operantes a falta de acuerdo expreso, y decidido que precisamente esas normas son las que ellos desean que rijan su acuerdo, y pueden haber decidido de este modo no variarlas por medio de ninguna estipulación expresa. No cabe duda de que hay veces en que esto ocurre en relación con normas legales particulares, especialmente en contratos comerciales. Pero ciertamente esto no siempre ocurre cada vez que las partes no han variado en forma expresa el efecto de estas normas. Es evidente que el derecho clásico estaba equivocado al tratar esas normas como basadas en la intención (presunta) de las partes solamente porque no había acuerdo expreso en relación con los asuntos tratados por estas normas. Desafortunadamente aún vivimos con los resultados de ese error.

Así también, el énfasis en la intención y en los acuerdos "implícitos" a menudo confundió a las cortes y a los juristas, quienes no pudieron percatarse de que una buena parte del derecho de los contratos tenía relación con obligaciones que emanaban de lo que las partes hacían, y no sólo de lo que las partes acordaban o prometían.

Y al imponer obligaciones en razón de lo que las partes habían hecho en lugar de lo que había sido su intención, necesariamente los jueces estaban utilizando sus ideas de equidad y justicia.

Ya se dijo anteriormente que la segunda idea abrazada por el concepto de libertad contractual fue que los contratos eran el resultado de una libre elección, y esto requiere ahora algún desarrollo. Una vez más vemos la influencia de las teorías políticas y económicas en marcha, porque la libertad de elección en la que las cortes estaban pensando, era una libertad en un sentido más bien restringido. Era libertad en el sentido de que nadie estaba obligado a ser parte de un contrato si no elegía serlo, libertad en el sentido de que en una sociedad competitiva todos tenían la opción de elegir con quien contratar, y libertad en el sentido de que las personas virtualmente pueden realizar cualquier contrato y en los términos que ellas elijan. La primera de estas proposiciones fue ciertamente muy verdadera en materias estrictas de derecho, y quizás también, en el siglo diecinueve en un sentido más amplio . Por cierto sólo en algunos pocos casos se dio que una persona estaba obligada legalmente a celebrar un contrato; virtualmente el único ejemplo de este tipo de obligación fue, de hecho, el de una persona realizando un "llamado público", así como el posadero y la empresa de transporte público quienes estaban (sujetos a ciertas salvaguardas) legalmente obligados a contratar con cualquier miembro del público que requiriera esos servicios. Así también en el siglo diecinueve, la competencia industrial y comercial generalmente otorgó una real libertad de elección, y fue sólo la llegada de los ferrocarriles y de los servicios de utilidad pública la que presagió la posterior tendencia a los vastos monopolios que redujeron la competencia en tantos campos. Pero la libertad en su tercer sentido, es decir la libertad para realizar cualquier tipo de contrato, en los términos que deseen, fue, incluso en el siglo diecinueve, de algún modo restringida. Siempre existía el dominante interés público o ser considerado - " política pública " como es llamada en el derecho de los contratos - y en ciertos casos las cortes retuvieron el poder para declarar la invalidez de algunos contratos por ser contrarios a la "política pública". Entonces, en una época relativamente temprana, la interferencia legislativa con la libertad contractual comenzó a jugar un papel más importante. Por ejemplo, ya en 1831 el

primero de las modernas *Truck Acts* comenzó a proteger a los empleados de la práctica del pago en especias en vez de dinero, y en 1845 el *Gaming Act* estipuló que las apuestas no serían exigibles como contratos. Sin embargo cuando uno considera la gran cantidad de legislación existente durante el presente siglo que restringe la libertad en este sentido, uno se inclina a concordar con los jueces del siglo diecinueve en el sentido que en dicha época existía una real libertad contractual.

Otra consecuencia del acento en la libertad contractual en el derecho clásico requiere de énfasis. Dado el dogma de que las obligaciones podían ser creadas por la intención de las partes, comenzó a ser un principio fundamental del derecho que los contratos fueran obligatorios y su cumplimineto posible de ser exigido tan pronto como fueren "hechos", es decir, tan pronto como fueren celebrados. Es de este modo absolutamente fundamental para la teoría clásica de los contratos el distinguir que *hacer* un contrato es algo distinto de *cumplir* un contrato. No era necesario demostrar ningún acto de ejecución o confianza para justificar una demanda basada en un contrato; tampoco era necesario demostrar de que el demandante había sufrido alguna pérdida con la ruptura del contrato. Era suficiente con que hubiera esperado alguna ganancia si el contrato era cumplido. Tenía derecho a recobrar daños por estas "expectativas perdidas ".

Incluso desde un comienzo el concepto clásico de libertad contractual sufrió de ciertas debilidades, y en el siglo veinte estas debilidades se magnificaron enormemente. Como hemos visto, el concepto clásico de libertad contractual abrazaba dos ideas, la de los contratos como basados en acuerdos de voluntades, y la de los contratos como un resultado de la libre elección. Es dudoso si alguna vez estas ideas fuero completamente válidassy ciertamente en la última mitad del siglo diecinueve los cambios en las condiciones sociales y económicas y en la práctica de las cortes ya indicaban que la libertad contractual se estaba transformando en una cosa bastante distinta de lo que se pretendía.

El desarrollo del derecho mismo durante la última mitad del siglo diecinueve afectó profundamente la importancia atribuida al acuerdo y a la intención de las partes. Siguió siendo verdadero, por supuesto, que los contratos eran usualmente creados como resultado de un acuerdo, aunque, como hemos visto, el enfoque objetivo a las cuestiones relativas a los acuerdos y la intención gradualmente se intensificó. Pero mucho más importante que esto fue el hecho de que a medida que fueron siendo más elaboradas las normas relativas a contratos, la absoluta complejidad del derecho creció de tal manera que de la teoría que

establecía que la mayoría de las normas estaban basadas en las intenciones de las partes se fue convirtiendo cada vez más en una ficción ilusoria.

En forma similar, con el surgimiento de detalladas leyes que regulaban contratos más específicos, como las ventas, agencia comercial, seguros, etc., el derecho tendió a volverse más estandardizado. Nuevamente el resultado consistió en que la real intención de las partes a menudo comenzó a ser menos importante. El derecho relativo a la venta de mercaderías puede ser tomado como un ejemplo de este proceso. A lo largo del siglo diecinueve, las obligaciones del comprador y del vendedor en los contratos de venta fueron detalladas en forma casi absoluta, usualmente bajo el fundamento de que estas obligaciones se basaban en la intención presunta de las partes. A medida que el derecho se fue desarrollando, las obligaciones se fueron estandardizando cada vez más, hasta que en 1893 fue posible codificar el derecho relativo a la venta de bienes en el *Sale of Goods Act de 1893*, ahora reemplazada por la ley de 1979. El resultado es que en la actualidad las obligaciones de las partes en ese tipo de contrato se pueden encontrar en el *Sale of Goods Act* y ya no es necesario atribuir estas obligaciones a intención alguna, presumida o de otro tipo. Es cierto que la intención de las partes aún tiene importancia en el sentido de que puede por mención expresa de las partes hacer variar o excluir las obligaciones que constan en la ley. Pero aún esta libertad de elección es ampliamente restringida en la actualidad a beneficio de los consumidores y de otros que se enfrentan con "términos tipo de comercio por escrito".

Al considerar la libertad contractual en su otro sentido, como libertad de elección, nuevamente encontramos debilidades en el derecho clásico. Por largo tiempo ha sido costumbre indicar dos debilidades en particular en el derecho clásico. En primer lugar, el derecho clásico de los contratos prestó poca atención a las desigualdades existentes entre las partes contratantes. Libertad contractual significaba que uno podía elegir con quien contratar, y se podía llegar a los términos que se quisiesen por acuerdo mutuo. Incluso en el siglo diecinueve esto fue sólo verdadero en un sentido restringido, esto es, era cierto si uno asumía que el poder negociador de todas las partes contratantes era similar, y este era un punto que el derecho clásico había aceptado ampliamente. Por supuesto, hubo casos como el de las personas incapaces, dementes, etc., en que se necesitaban estipulaciones especiales, pero en general el derecho asumió que cada persona podía valerse por si misma, y si una persona entraba en un contrato gravoso o severo sólo podía culparse a sí mismo ya que existía libertad contractual y podría haber contratado con quien quisiera. Claramente esta aseveración era falsa aún

en el siglo diecinueve. La fuerza negociadora del empleador, por ejemplo, era normalmente mayor que el de sus empleados cuando los sindicatos aún estaban prohibidos en su conjunto o eran incipientes. Los pequeños comerciantes, una vez más, no eran rivales para la gran compañía ferroviaria, etc. y etc.

Una razón por la que el derecho tradicionalmente prestó poca atención a la desigualdad de poder negociador se debió a que estas desigualdades eran concebidas como materias que implicaban justicia *distributiva* más que justicia *conmutativa*. El derecho de las obligaciones tradicionalmente sólo se ocupó de materias relativas a la justicia *conmutativa*, es decir, poniendo en orden las cosas que no habían andado bien producto de un incumplimiento del contrato, o por un perjuicio causado por una persona a otra. Se pensaba que las desigualdades en la sociedad que resultaban de la forma en que se distribuían los recursos y la riqueza, eran materias esencialmente políticas, que debían ser abordadas por el Parlamento, e incluso el Parlamento no tenía excesivo interés en la redistribución de la riqueza en la Inglaterra del siglo diecinueve.

Esta falta de preocupación por cuestiones distributivas estaba relacionada con otra característica del derecho clásico, esto es, su naturaleza más general y abstracta. Fue, en efecto, durante el período clásico que el concepto mismo de *derecho general* de contratos fue fundamentalmente desarrollado. Previamente habían habido cuerpos separados de leyes relacionados con diferentes tipos de contratos, arriendos, compraventa de mercaderías, seguro, etc. Ahora la tendencia fue a generalizar a partir de estos tipos de contratos, y esta tendencia bien puede haber ayudado al derecho a desarrollar su enfoque neutral a problemas distributivos. Si el derecho debía ser el mismo para todo tipo de contratos, para compraventas, arriendos, promesas de matrimonio, etc. es más difícil entonces desarrollar reglas especiales para proteger tipos particulares de personas, como arrendatarios o empleados.

El segundo problema en este aspecto de la libertad contractual fue que el derecho clásico tomó poco en cuenta las presiones económicas y sociales que en muchos casos podía virtualmente forzar a una persona a celebrar un contrato. Incluso en el siglo diecinueve esto fue verdadero en muchas formas. La mayoría de las personas debían buscar una forma de ganarse la vida, y a menudo esto significaba celebrar un contrato de trabajo. Los suministros de agua y gas se hicieron rápidamente esenciales para las clases acomodadas, y era improbable obtenerlos de más de una fuente en una zona en particular. Los ferrocarriles se estaban transformando en la forma moderna de viajar, e incluso a fines del siglo

diecinueve era la única forma de recorrer distancias largas por tierra. Por cierto, en esos momentos no existía ni elección ni competencia.

Esta debilidad evidente en la teoría clásica fue más importante con posterioridad, cuando los monopolios y las prácticas restrictivas se fueron extendiendo; pero en los primeros tres cuartos del siglo diecinueve, durante el apogeo del derecho clásico de los contratos, la economía británica era altamente competitiva, y esto significó que había un alto grado de libertad de elección en el mercado, tanto en la teoría como en la práctica.

4. LA DECADENCIA DE LA LIBERTAD CONTRACTUAL 1870-1980

Durante aproximadamente los últimos cien años los cambios en el pensamiento político, en las condiciones sociales y económicas, y en el derecho, han estado teniendo lugar a un paso siempre creciente. Después de la época de la libertad contractual (quizás más exactamente ubicada entre 1770 y 1870) no es difícil de identificar el período entre 1870 y 1980 como un período de decadencia gradual de la creencia en la libertad contractual. Los desarrollos durante este período representaron en cierto grado un retorno hacia antiguas tradiciones anteriores a la época del individualismo y la libertad contractual. Por ejemplo, la ideología paternalista que comenzó a influenciar gran parte del derecho entre (digamos) 1950 y 1980 fue en muchos aspectos más cercana a aquella del siglo dieciocho que a la del siglo diecinueve. Gran parte de este cambio fue influenciado por una creencia generalizada de que el derecho contractual clásico ya no se ajustaba, en muchas situaciones, a los acontecimientos del mundo moderno. Sin duda que en el sector meramente comercial, en donde los comerciantes contrataban unos con otros para la compra y venta de mercaderías, o para la construcción o manufactura de plantas y edificios, gran parte de la libertad contractual, aún en su sentido clásico, todavía seguía existiendo. Pero incluso aquí el derecho fue altamente modificado, y hacia 1980 el derecho contractual clásico parecía estar desmoronándose rápidamente. Tres factores específicos deben ser mencionados por haber tenido la mayor influencia en ayudar a la destrucción de la coherencia del derecho contractual clásico. El primero fue la aparición y el uso generalizado del contrato tipo, el segundo fue la decreciente importancia asignada a la libre elección e intención como fundamentos de la obligación legal, y el tercero fue el surgimiento del

consumidor como parte contratante (y quizás aún más como litigante). Estos tres factores están, por supuesto, interrelacionados.

Contratos tipo.

En una enorme cantidad de casos, la idea de que los contratos se basan en un acuerdo, sólo es verdadera en un sentido muy restringido. Si bien era cierto que, en la mayoría de los casos, la verdadera creación de un contrato (y por tanto los términos esenciales) requería del acuerdo de las partes, no era ya verdadero (si es que realmente lo fue alguna vez), que en muchas circunstancias los términos específicos de un contrato dependieran del acuerdo de las partes. Una razón para esto fue el desarrollo del contrato tipo. Hoy la mayoría de los contratos no se componen de términos individualmente negociados o hechos a la medida. Comenzaron a ser hechos en base a formas estándar impresas ofrecidos por grandes organizaciones, comúnmente bajo el principio de "tómalo o déjalo". El consumidor sigue siendo "libre" en teoría pero su elección puede bien estar restringida a un "tomarlo" o "dejarlo". Así, pese a que nadie puede obligar a un pasajero a viajar en tren, si así lo quisiera tendría que hacerlo en los términos y condiciones impuestas por las compañías. No podría negociar sus propios términos. De forma similar, si un dueño de casa quisiera obtener suministros de gas o electricidad, tendría que celebrar un contrato en los términos fijados por los suministradores. Incluso cuando los bienes y servicios en cuestión no eran monopolizados por una sola organización, era frecuente encontrar que los términos y condiciones ofrecidas al público eran prácticamente, si no exactamente, idénticas.

Hacia mediados del siglo veinte estos contratos tipo habían llegado a ser uno de los mayores problemas del derecho contractual. Podían ser hallados en cada aspecto de la vida. En la mayoría de los casos sólo era verdadero decir en un sentido muy restringido que el contrato era el resultado de un acuerdo. Frecuentemente no existía ninguna oportunidad, y por lo tanto no verdadera libertad, para negociar términos específicos en estos casos. Los términos eran impuestos por una parte, y la otra no tenía ninguna posibilidad más que aceptarlos o no contratar. Precisamente por la naturaleza misma de la situación, estos términos tendían a ser mucho más favorables para la organización que suministraba los bienes y servicios en cuestión que para el individuo que los recibía. La organización tenía todas las ventajas por sobre el individuo. Normalmente tenía la ventaja de grandes recursos, de la mejor asesoría legal, de ser capaces de litigar, si llegaba a ser necesario, sin tener que preocuparse

excesivamente por el costo, y , por supuesto, de saber que el individuo, aunque se retorciere de rabia, no podía realmente prescindir de sus servicios. Un rasgo extremadamente común y fastidioso de los contratos tipo era la presencia de una "cláusula de exención", que muchas veces permitía que la organización no fuera responsable virtualmente bajo ninguna circunstancia.

Los contratos tipo también llegaron a ser muy usados en las transacciones entre comerciantes. Los acuerdos comerciales para la venta de mercaderías eran normalmente registrados en "formularios" o "recibos" con cláusulas impresas; los contratos para el transporte marítimo de bienes fueron usualmente registrados en una póliza de flete; los contratos de seguro eran registrados casi siempre en pólizas impresas, etc... En situaciones de esta naturaleza pudo haber espacio para la negociación, pero el caso es que la mayoría de las cláusulas en semejantes contratos no eran negociadas sino impuestas.

La totalidad de la estructura teórica de la libertad de contratar comenzó a parecer cuestionable a medida que la magnitud de este problema era percibida. Se hizo cada vez más claro que la gran mayoría de los principales contratos celebrados por el ciudadano medio eran realizados en términos que en mayor o menor medida le eran impuestos. Por ejemplo, la casa en que vivía probablemente estaba hipotecada en favor de una sociedad constructora en los términos y con la tasa de interés fijados por la propia sociedad, generalmente en acuerdo con todas las otras sociedades constructoras. Sin duda que pudo haber sido capaz de conseguir una hipoteca en alguna otra parte, pero habría encontrado pequeñas diferencias en los términos, y prácticamente ninguna en la tasa de interés. De otro modo, la casa le pudo ser alquilada por la autoridad local, en cuyo caso los términos del arrendamiento eran enteramente determinados por la autoridad, y estaban ahí para ser aceptados, no para ser negociados (6). Los suministros de gas y electricidad se realizaban, como ya se mencionó, en los términos fijados por los suministradores. El amoblado o el automóvil comprados a plazo por el consumidor eran siempre el objeto de contratos cuidadosamente redactados por corporaciones financieras para salvaguardar sus intereses, y el ciudadano medio estaría sumamente sorprendido si pudiera comprender estos contratos, y menos aún los habría convenido de manera alguna. La póliza de seguro que él tomaba por su vida o su automóvil, estaba también en un formulario estandarizado, obviamente debiendo ser ajustadas las primas para coincidir con el status del individuo interesado. Aún cuando se dirigía a su oficina o fábrica el ciudadano medio era perseguido por los contratos tipo. Por ejemplo, en el bus o tren su boleto era emitido (y por cierto, aún es), sujeto a innumerables términos

que el ciudadano podía examinar en la oficina de la organización pertinente, si es que se interesaba en leer cosas incomprensibles para él. En muchos casos, también, su empleo era regulado por un contrato tipo. Esta es una lista que podría continuar en forma más o menos indefinida, pero se han dado suficientes ejemplos para demostrar la magnitud del problema.

Todo esto no significa que hubiese algo malo con los contratos tipo en si mismos. Tenían las ventajas de ahorrar tiempo, problemas y gastos en la negociación de los términos. También tenían la ventaja de que un fallo para un caso podría muy probablemente servir de guía para problemas disputados en otros casos. En una era de producción en masa es apenas necesario enfatizar las ventajas del artículo hecho en serie, y esto se aplica a los contratos en serie como a cualquier otra cosa. Pero lo que estaba absolutamente claro era que muchos términos de un contrato tipo característico no eran *convenidos* en ningún sentido real. La realidad era que el típico contrato era hecho en base a un conjunto de términos estándar o fijos con unos pocos espacios en blanco para la inserción del nombre de las partes, la materia específica del contrato, el precio, y quizás uno o dos detalles más. Esto por supuesto podría variar dependiendo del caso, y probablemente podría existir alguna posibilidad para negociar algunos de estos espacios en blanco, pero el resto de los términos ya estaban fijos, a menudo en formularios impresos. Algunas veces eran confeccionados por una parte, entonces posiblemente representaban sólo sus propias intenciones; otras veces no eran confeccionados por ninguna de las partes, por ejemplo, los contratos tipo de construcción e ingeniería eran (y son) confeccionados por el *Royal Institute of British Architects* y el *Institute of Civil Engineers*, caso en el cual, ambas partes tendrían sólo las más vagas "intenciones" respecto de lo que los términos de su contrato realmente disponen. A menudo, estos formularios tipo eran usados a lo largo de toda una industria, siendo, en lo esencial, formas contractuales ampliamente acordadas, a las cuales se adherían todas las firmas, privando así a los consumidores de los reales beneficios de la competencia y la libre elección.

La decadencia del rol de la libre elección.

Como hemos visto, incluso durante la última parte del siglo diecinueve la importancia asociada a la intención de las partes ya estaba disminuyendo, existiendo en el derecho razones técnicas para ello. Una de esas razones era el simple hecho de que el derecho estaba ganando en complejidad. Este proceso también, continuó a lo largo de este siglo y en cierto grado aún continúa hoy. Por ejemplo, en años recientes, a medida que las cortes han distinguido entre los

diferentes tipos de daños que pueden ser indemnizables, y han refinado y modificado las reglas que rigen el incumplimiento de los contratos, el derecho relativo a las acciones por incumplimiento de contrato ha pasado a ser mucho más desarrollado. Se hizo cada vez más difícil decir, como la clásica teoría de los contratos sugería, que la única función de las cortes en el derecho contractual era "hacer cumplir" los contratos. Cada vez resultaba más obvio que raramente las cortes "hacían cumplir" los contratos en un sentido literal - lo que hacían era imponer alguna clase de sanción por la violación de los contratos, en particular, con una indemnización de perjuicios, y también se hizo más obvio que la elección de la sanción apropiada era un acto de poder imaginativo ejercido por las cortes.

Una segunda razón para la disminución en la creencia en el valor de la libertad de elección fue el creciente escepticismo respecto de la libre elección en el mercado. Como hemos visto, la desigualdad de poder negociador, las presiones sociales y económicas (a menudo las presiones de la pobreza), y el uso de formularios tipo hacían pensar ahora que, en muchas situaciones, no existía una real libertad de elección, cualquiera que fuera la teoría del mercado. Con semejante enfoque sobre las desigualdades de poder negociador, y sobre las presiones sociales y económicas como debilidades del concepto clásico de libertad contractual (y en cierto grado, quizás también, sobre los contratos tipo), bien puede ser que los críticos estuvieran atacando los blancos equivocados. Es verdad que la libertad de elección era menos real y menos amplia que lo que la teoría clásica asumía, especialmente en el período que comenzó alrededor de 1870. Pero el verdadero problema aquí no era la desigualdad de poder negociador en sí misma, ni las presiones sociales y económicas de la pobreza; el verdadero problema era el crecimiento de los monopolios y de las prácticas restrictivas de todo tipo. Como veremos más acabadamente en el capítulo 17, alrededor de 1870 y 1950 la economía británica se cubrió con una vasta red de prácticas restrictivas y monopolios. Fueron estos acontecimientos los que restringieron la libertad de elección, y llevaron a muchos críticos y reformadores a dudar respecto de la eficacia y justicia de la libertad contractual. Pero los críticos malentendieron las causas de las dificultades. A pesar de que desde la década de 1890 en adelante, los economistas ya habían advertido la amenaza del incremento de los monopolios y las prácticas restrictivas, los políticos, el mundo de los negocios y la opinión pública no eran hostiles del todo a estos acontecimientos. Por cierto, a menudo eran vistos como medidas de protección necesarias diseñadas para permitir a la industria británica hacer presente a la amenaza de la competencia del resto del mundo industrial. Ha sido sólo en la

última década o dos que a comenzado a ser más claramente entendido que si el mercado es verdaderamente competitivo, y si el monopolio y las prácticas restrictivas están bajo un fuerte control legal, los asuntos como la desigualdad de poder negociador, presiones en el mercado, y el uso de formularios tipo no constituyen necesariamente amenazas serias para la libertad de elección.

Pero probablemente la principal razón del descenso de la creencia en la libertad de elección fue simplemente un cambio en los valores políticos. El período entre 1870 y 1980 fue un período en que los valores colectivistas, e incluso los socialistas, se extendieron en Inglaterra. La idea de que el intercambio libre y voluntario era la clave para la prosperidad económica, y por cierto, quizás para una sociedad más libre y más satisfecha, cayó en abrupto descenso durante este siglo. En primer término, existió una crecida sofisticación en el entendimiento económico de las limitaciones del sistema de libre contrato. En particular, el reconocimiento del problema de las "externalidades" fue una gran ganancia para el entendimiento. Una externalidad es, *grosso modo*, un efecto colateral de un intercambio libre que afecta a terceros, y pudo observarse que incluso si un intercambio es beneficioso para las dos partes que lo realizan, no será del interés público si hay externalidades que exceden la ganancia privada. La Inglaterra del siglo diecinueve vio surgir problemas masivos de externalidades a partir de la Revolución Industrial - ciudades sucias, viviendas insalubres, enfermedad y contaminación, etc, pueden todos ser vistos como costos externos impuestos por la industria a los terceros, esto es, el público. Frecuentemente estas externalidades eran el resultado de contratos privados, por ejemplo, para el alquiler de cuartos sin desagües que las partes libremente convenían, y que sin lugar a dudas querían realizar. Ahora era cada vez más claro que la sociedad estaba perfectamente habilitada para prohibir semejantes contratos, si es que era necesario, por medio de la construcción de desagües y forzando a los residentes a pagar por ella.

Pero incluso además de las limitaciones económicas del sistema de libre contrato, también llegó a ser un amplio sentir que semejante sistema conducía a resultados inaceptables e injustos. El débil y el pobre, el vulnerable y el explotado, se sentían en necesidad de protección por el Derecho. Cada vez era más dominante el pensamiento de que si eran dejados solos para que hicieran sus propios contratos, serían inevitablemente aventajados por las otras partes contratantes ricas y poderosas. Así, el Derecho fue llamado en ayuda para interferir en los contratos en múltiples formas, prohibiendo cierto tipo de contratos, cierto tipo de términos contractuales, o insistiendo que otros contratos debían

conferir para una parte u otra derechos que no estaban realmente contenidos en el contrato.

Por estas razones y por otras similares comenzó a parecer que grandes áreas de actividad no podían ser dejadas a la libre elección de las partes contratantes, y que el derecho contractual era, de todas maneras, menos una materia de libre elección como se había pensado antes. Vastas áreas del derecho llegaron a ser reguladas por la legislación, gran parte de la cual obstruyó o anuló totalmente la libertad contractual. Así también algunos jueces comenzaron a admitir que a veces las soluciones contractuales eran "impuestas" sobre las partes; y existían también muchas circunstancias en las cuales actos que antiguamente se decía que dependían de la intención de las partes, venían ahora a ser tratados como si dependieran simplemente de reglas de derecho. Por ejemplo, la libertad de las partes para elegir qué ley gobernaría su contrato no se consideraba tan propicia como antes; la operación de la "doctrina de la frustración" fue extendida y liberada de su ficticia dependencia de las intenciones de las partes; la libertad de las partes para declarar por anticipado cuales habrían de ser los resultados de un incumplimiento del contrato fue limitada por la doctrina del "incumplimiento fundamental", pese a que ella fue a la larga abandonada después de que la legislatura tratara el problema en forma más comprensiva de lo que las cortes jamás pudieron.

Por otra parte, habían muchas otras materias, como los problemas de interpretación o integración del contrato que mientras aún se "decía" que dependían de la intención de las partes, eran frecuentemente decididas sin hacer ningún intento real por indagar esta intención. Un área clave del derecho era la "doctrina de la frustración", la cual, como hemos visto, habilita a una corte para declarar que un contrato ha terminado cuando sin culpa de cualquiera de las partes se ha hecho imposible de ejecutar. En la teoría clásica, era tradicional explicar la doctrina como descansando en un término implícito en el contrato original, y por lo tanto, en la intención de las partes, pero está teoría fue abandonada más tarde. El caso seminal fue la decisión de la Cámara de los Lores en *Davis Contractors v. Fareham UDC* (7) y el siguiente pasaje del discurso de Lord Radcliffe en este caso explica como fue posible moverse de una teoría a la otra:

Lord Loreburn atribuye la disolución a un término implícito del contrato que fue realmente hecho. Este enfoque está en conformidad con la tendencia de las Cortes Inglesas de referir todas las consecuencias de un contrato a la voluntad de aquellos que lo celebraron. Pero hay alguna dificultad lógica en ver como las partes pueden incluso tácitamente haber prevenido algo que, *ex*

hypothesi, ellos jamás esperaron ni previeron, y la adscripción de la frustración a un término implícito del contrato ha sido criticada por oscurecer la verdadera acción de la corte, la que consiste en aplicar una norma objetiva del derecho contractual a las obligaciones contractuales que las partes se han impuesto sobre si mismas... Pero aún puede ser de alguna importancia recordar que, si ha de aproximarse a la materia por la vía del término implícito, la solución de cualquier caso particular no será hallada inquiriendo por lo que las partes mismas habrían acordado de haber, como no lo hicieron, estando prevenidas. No se trata sólo de que ninguno pueda responder a esa pregunta hipotética: sino también de que la decisión debe darse "prescindiendo de los individuos interesados, sus temperamentos y defectos, sus intereses y circunstancias". (8) El efecto legal de la frustración "no depende de su intención o de su opinión, o incluso de su conocimiento del evento". (9) Por el contrario, parece ser que, cuando el evento ocurre, "el significado del contrato debe ser, no lo que las partes quisieron (porque ellas jamás tuvieron pensamiento ni intención respecto del evento), pero si aquello que las partes, como hombres honrados y razonables, hubieran presumiblemente acordado entonces si, teniendo semejantes posibilidades en vista, ellos hubieran hecho provisión expresa respecto de sus derechos y obligaciones en el evento de su ocurrencia".(10)

A estas alturas, parecería que las partes se han convertido de tal modo en espíritus sin cuerpo, que a sus personas reales les debería ser permitido descansar en paz. En su lugar se alza la figura del hombre honrado y razonable. Y el vocero del hombre honrado y razonable, que representa, después de todo, no más que el concepto antropomórfico de justicia, es, y debe ser, la propia corte. (11)

Las palabras de Lord Radcliffe pueden ser parafraseadas como sigue. No es el acuerdo propiamente tal o la intención de las partes lo que importa, pero si aquello que consideran, como partes razonables, haber acordado o querido. Lo que necesitamos saber, por lo tanto, es lo que el hombre razonable habría acordado, o querido, en estas circunstancias. Pero en el hecho, es la corte la que representa al hombre razonable, y es así como la corte tiene que expresar lo que ella piensa que las partes debieron razonablemente haber acordado o querido. Ahora queda claro que la solución que finalmente impone la corte no es dependiente en ningún sentido real de la intención de las partes, sino que está simplemente basada en una norma de derecho que es como cualquier otra norma de derecho.

Volviendo a los desarrollos que afectan el concepto de libertad contractual en el sentido de libertad de elección, de nuevo encontramos que hacia mediados del siglo veinte se creía cada vez más que el derecho clásico de los contratos era anticuado. Aún era cierto que, sin lugar a dudas, habían pocos casos en los que

una persona estaba bajo una obligación legal de celebrar un contrato. Pero presiones de un tipo u otro, a las cuales el derecho común daba poca relevancia tenía escasa noticia, podían ser tan importantes como una obligación legal, y estas presiones virtualmente podían forzar a una persona a celebrar un contrato que ella no tenía un gran deseo de realizar. En algún grado esto ya era efectivo en el siglo diecinueve, y se hizo infinitamente más efectivo posteriormente. Por ejemplo, una persona podía ser compelida a unirse a un gremio (así celebrando un contrato) a fin de ejercer su oficio y ganarse la vida. O también, un minorista podía estar virtualmente obligado a contratar con un mayorista para obtener los bienes necesarios para su negocio. De modo similar, cada dueño de casa, como hemos visto, está prácticamente obligado a contratar con los suministradores de gas y electricidad.

También comenzarán a tener lugar otros cambios con los que el derecho de los contratos tuvo que tratar. Por ejemplo, una persona puede tener un derecho establecido por ley para ser abastecido con bienes o servicios en ciertas circunstancias (por ejemplo bajo el Servicio de Salud), y entonces los problemas pueden surgir por la naturaleza e incidencia de la relación entre las partes. También está teniendo lugar un desarrollo similar en otras áreas que pueden crecer en importancia a todo nivel, si el resurgimiento de los principios clásicos lo permite. Por ejemplo, un arrendatario de locales comerciales tiene, sujeto a ciertas condiciones, un derecho establecido por ley a un nuevo arriendo cuando expire el antiguo. El propietario está "legalmente obligado" a conceder el nuevo arriendo. Para el tradicional abogado especializado en derecho contractual, una obligación de celebrar un contrato parece una contradicción en los términos, porque los contratos se han visto tradicionalmente como el resultado de la libre elección, aunque este ejemplo es sólo uno de muchos que existen en el derecho moderno. Quizás un ejemplo aún más sorprendente, ha sido la legislación "derecho a comprar" (contenido en el *Housing Act 1985*) imponiendo una obligación legal sobre los municipios para vender sus casas a arrendatarios municipales que cumplen con ciertas condiciones simples. Es bien sabido que muchos municipios objetaron fuertemente esta legislación, pero no obstante es derecho y ha sido comúnmente ejecutado contra ellos.

También existen leyes modernas dirigidas a la prevención de la discriminación que claramente se basan en la suposición de que una parte "no debería" ser libre para rechazar el celebrar un contrato bajo ciertas condiciones. Por ejemplo, la *Race Relations Act de 1976* restringe la libertad de diversas personas tales como empleados y dueños de almacén, para rechazar el trato con

otro a causa de sus orígenes raciales. En forma similar la *Sex Discrimination Act de 1975* hace ilegal para los proveedores de variados servicios discriminar a las mujeres (o a los hombres). Ahora estas leyes otorgan, en ciertas circunstancias, una acción civil compensatoria a favor de la persona que ha sido discriminada.; en efecto, el demandante puede recobrar una compensación de una persona por negarse ilegalmente a celebrar un contrato. Lo curioso es que los abogados generalmente no perciben tales leyes como afectando los principios del propio derecho contractual: allí, la libre elección sigue siendo, en teoría, el último fundamento del derecho.

También existen muchos tipos de intercambios económicos que fueron eliminados en bloque del área del derecho contractual y tratados por el derecho público. De este modo, las autoridades gubernamentales y locales dotaron al consumidor con una amplia gama de servicios, tales como educación gratuita para sus hijos, servicios médicos gratuitos en el Servicio Nacional de Salud, recolección de basura gratuita desde sus casa, etc... Estos servicios "gratuitos" obviamente no son gratis - sino que son pagados de una u otra forma por los impuestos y no por medio de un contrato. Aquí no existe libre elección del todo - el consumidor debe pagar una carga o un impuesto fijados por la legislación o las autoridades locales, y a menudo tiene que pagar ya sea que necesite o no el servicio, y ya sea que el servicio sea otorgado o no; y lo que es más, no tiene compensación si el servicio es pobre o ineficiente. El dueño de casa a quien no se le recogen los tarros para la basura no puede negarse a pagar alguna parte de sus contribuciones, ni tampoco puede demandar por incumplimiento de contrato. Los abogados no conciben del todo estos intercambios como materias del derecho contractual, porque son regidos por principios legales bastante diferentes. Se ha dicho, por ejemplo, que cuando una persona obtiene una receta de un químico en el Servicio Nacional de Salud no está celebrando un contrato, pero está ejerciendo un derecho establecido por ley. Del mismo modo, la relación existente entre un médico que trabaja en el Servicio de Salud y la autoridad sanitaria local esté en gran parte regulada por complejas condiciones legales, y es incierto si es que queda algún espacio para la aplicación de los principios comunes del derecho de contrato (12). Pero pese a que estos intercambios pueden no ser contratos, ellos realizan una función económica similar, y mientras más amplios sean esos intercambios, menos amplio será el campo del contrato y en consecuencia del derecho contractual. El derecho de los contratos puede permanecer inalterado, pero se aplica cada vez a menos transacciones, y este fue

uno de los giros observables en nuestro derecho tal vez especialmente en el período que va aproximadamente entre 1945 y 1980.

El descenso de la importancia asociada a la libre elección e intención como fuentes de la obligación también pareció llevar a una serie de cambios fundamentales en el derecho y el pensamiento legal, aunque estos cambios no eran siempre obvios a simple vista. Habían, por ejemplo, crecientes signos de una voluntad favorable de las cortes "para hacer un contrato para las partes". Cuando dos partes entran en ciertas relaciones (por ejemplo, una mujer y un hombre instalan una casa en conjunto), puede suceder que con posterioridad ocurran disputas, y que una parte recurra a las cortes en busca de ayuda. En semejantes casos las cortes comenzaron a demostrar una considerable capacidad para "hacer contratos". Otro gran cambio en el derecho pareció descansar en una creciente renuencia a reforzar los contratos "totalmente ejecutorios"* con una sentencia por daños exclusivamente provenientes de la "pérdida de expectativas", excepto quizás en obvios casos comerciales. Como vimos anteriormente, los daños por lucro cesante son sólo justificables si es aceptado que una persona puede imponer una obligación sobre sí misma, a lo menos en principio, solamente por proponérselo. Hasta 1980 hubo algunas señales de que las cortes se sentían disconformes con algunas obligaciones de este tipo. Por otra parte, cuando se causaban pérdidas reales a una persona por haber razonablemente confiado en lo que otra había dicho o hecho y cambiado su posición con un resultado perjudicial, la tendencia a protegerlo con una sentencia de "daños por confianza" fue más fuerte. Retornaremos a este tema en el Capítulo 22. Así también, hubo una pronunciada tendencia de las cortes a delinear distinciones entre los contratos ejecutorios - no ejecutados o que no habían generado actuaciones de la otra parte basados en ella - por un lado y contratos ejecutados en parte, por el otro.

La decadencia de la importancia del derecho contractual fue, en cierta extensión, correspondida por un crecimiento en importancia de otras áreas del derecho de las obligaciones. En particular, hubo un gran crecimiento en la importancia del derecho de los daños extracontractuales, por un lado, y las acciones restitutorias por otro. Estas otras dos áreas del derecho de obligaciones en cierto sentido lindan y se superponen con amplias áreas del derecho de contratos. El derecho de los daños extracontractuales a menudo protege a una parte - dándole un derecho a indemnización de perjuicios - la cual ha confiado

* Los *executory contracts* del Derecho Anglosajón corresponden a aquellos contratos que no han sido aún completamente perfeccionados o cumplidos. Son contratos cuyas obligaciones suponen un cumplimiento futuro (*n.t*).

razonablemente en la otra, incluso aunque no se hubiera celebrado ningún contrato entre las partes. Por cierto, hoy difícilmente es ir muy lejos el afirmar que existe un principio general en el derecho según el cual la razonable confianza en otro pueda crear obligaciones, incluso fuera del área de los contratos (13). En cierto grado esto es reconocer que el derecho puede imponer una obligación sobre una persona que se ha negado a asumir una obligación voluntariamente. En forma similar, las acciones restitutorias pueden imponer una obligación sobre una persona para pagar por beneficios recibidos aún cuando realmente no ha consentido o prometido pagar por esos beneficios. La expansión de las obligaciones es aquí, por lo tanto, equivalente a imponer obligaciones a partes que no las han asumido voluntariamente. Hasta ahora no existe un principio *general* en el derecho inglés que habilite a una persona a reclamar la restitución de beneficios generados sin contrato - por ejemplo, la Cámara de los Lores ha decidido recientemente que una persona que rescata el bote abandonado de otra, el cual está peligrosamente a la deriva, no puede reclamar una recompensa por salvamento si el rescate ocurre en un río, aunque por mucho tiempo ha existido ese derecho cuando un buque de mar es salvado en forma similar (14). Pero recientemente la Cámara de Los Lores ha hecho mucho para ampliar el espectro de las acciones restitutorias (15), pudiendo el desarrollo de esta forma de responsabilidad indicar un rol reducido para el derecho de contratos.

El crecimiento de la protección al consumidor.

Como vimos anteriormente, el derecho contractual clásico estuvo mayormente desligado de problemas de desigualdad de poder negociador. Pero en el siglo veinte, como el Parlamento comenzó en forma creciente a utilizar el sistema impositivo para redistribuir la riqueza, a menudo interfirió también en la libertad contractual para procurar proteger al débil del fuerte o del rico. Por ejemplo, la legislación laboral fue aprobada para conferir derechos a los trabajadores que no podían ser obtenidos por medio del contrato, a los trabajadores; la legislación de compra a plazo fue realizada para proteger a los arrendadores del trato incorrecto de las compañías financieras; y en forma similar se promulgó la legislación referente a los propietarios y arrendatarios para así conferir mejores derechos a estos últimos.

La legislación para la protección del consumidor también utilizó a menudo la creación de ofensas criminales para conductas que en el derecho común sólo serían un incumplimiento de contrato o incluso una acción perfectamente legal. Así, por ejemplo, la *Trade Descriptions Act de 1968* convirtió en infracción ofrecer

bienes para la venta en lenguaje falso o engañoso. Un incumplimiento de las disposiciones de tal estatuto puede tener su efecto en lo contractual, a pesar de que el estatuto está sólo relacionado directamente con el derecho penal. Por ejemplo, ahora es posible obtener una orden de compensación de una corte criminal que condena a una persona por una infracción que ha causado pérdidas o perjuicios a otra. Otros estatutos a veces obligan al vendedor de *mercaderías* a proporcionar al comprador una nota conteniendo ciertos detalles prescritos. El incumplimiento de tales disposiciones usualmente se convierte en una infracción con el resultado indirecto, una vez más, de que el contrato puede no ser exigible por la parte culpable.

La ley de protección al consumidor más importante relativa a los derechos contractuales es el *Unfair Contract Terms Act de 1977*. Se le da mayor profundidad al tratamiento de esta ley en el Capítulo 16, y aquí es suficiente decir que la Ley restringe ampliamente el uso de "cláusulas de exención" por medio de las cuales las partes contratantes se protegen de la responsabilidad legal. La ley se extiende más allá de la protección a los consumidores, desde que también opera, dentro de ciertos límites, para los hombres de negocio que contratan sobre "términos escritos estandarizados".

La interferencia legislativa en la libertad contractual no está, por supuesto, siempre dirigida a reparar el equilibrio entre el débil y el fuerte. Por ejemplo, gran parte de la interferencia legislativa en la libertad contractual está diseñada para fomentar vastos objetivos de interés público. Por muchos años, por ejemplo, ha sido política de sucesivos gobiernos británicos estimular el desarrollo de la agricultura para que la nación no tenga que ser tan dependiente de la importación de alimentos - algo que hace al país especialmente vulnerable en tiempos de guerra. En *Johnson vs.Moreton* (16) la Cámara de los Lores contaba con esta política pública para sostener que un arrendatario agrícola no podía contratar fuera de las protecciones conferidas a él por el *Agricultural Holdings Act* 1948. Por cierto, tal era la atmósfera general en ese tiempo que la Cámara de los Lores insistió que ya no se podía aceptar en forma general que los derechos conferidos por ley pudiesen negociarse libremente. El problema estaba en la interpretación del estatuto particular, el cual era realizado sin ninguna presunción de que la libertad contractual prevalecía sobre los derechos establecidos por ley.

Así también, se hizo más común en este período encontrar legislación con objetivos económicos públicos, como el control de la inflación. Tal legislación a menudo terminó limitando la libertad contractual, por ejemplo, estipulando que un depósito mínimo era necesario en ciertas transacciones de compra a plazo, o

incluso (como se hizo más de alguna vez) intentando controlar los incrementos de salario por ley. Una vez más, en un diferente tipo de casos, la cooperación internacional fue a menudo el origen de legislación que, en gran parte, codificó los términos de un contrato particular, por ejemplo, contratos para el transporte de bienes por mar.

Las cortes responsables del manejo del *common law* de los contratos rara vez intentaron ir por el mismo camino, esto es, fueron muy vacilantes en desarrollar principios que interfirieran en la libertad contractual y que anularan contratos libremente celebrados, incluso cuando una de las partes era notoriamente más débil que la otra e incapaz de proteger sus propios intereses. Pero a mediados del siglo veinte los jueces fueron claramente influenciados por el mismo tipo de consideraciones que inspiraron gran parte de esta clase de legislación - es decir, sintieron una simpatía por el pequeño consumidor, la parte contratante débil que se encontraba a sí mismo obligado a un contrato de empleo o de compra a plazo injusto o severo; y a pesar de que las cortes rara vez utilizaron el poder para anular los términos reales de un contrato podían, y a menudo trataron, de ayudar a la parte más débil. Por ejemplo, "incluyendo" términos apropiados, o por medio de un proceso benevolente de "construcción" de los términos que efectivamente eran hallados en el contrato. Hacia el tercer cuarto del presente siglo, escritores docentes, y posteriormente jueces y abogados en general, comenzaron a preocuparse de que la presencia de desigualdad de poder negociador socavaba la legitimidad de los contratos, y así acababa con uno de los soportes que antes se usó para justificar la imposición de la coerción estatal.

5. ACONTECIMIENTOS A PARTIR DE 1980: ¿RETORNO A LOS PRINCIPIOS CLÁSICOS ?

Durante la década pasada, el clima político y económico ha cambiado dramáticamente en Inglaterra, y las virtudes de los principios del libre mercado están siendo proclamados una vez más. De muy diferentes maneras las políticas del gobierno y del Parlamento han ido marcha atrás desde la elección de la Sra. Thatcher para su primera administración en 1979. Si, como ha sido sugerido más arriba, la decadencia del rol de la libre elección en el período entre 1870 y 1980 fue una de las causas del descenso de la importancia de la libertad contractual, entonces, debe reconocerse que en los últimos diez años arduos esfuerzos han sido realizados para revigorizar la libre elección. Al mismo tiempo ha habido un

considerable resurgimiento de la creencia en las virtudes económicas del sistema de libre mercado y también ha habido un creciente interés (especialmente en Estados Unidos de Norteamérica) en la relación entre los principios legales y económicos. Parte de ese trabajo ha llevado a una re-examinación de las normas legales e ideas que han dominado por cien años o más, y esto ha desafiado algunas de las asunciones que subyacían al descenso de la importancia de la libertad contractual. Todo esto sugiere que la decadencia de la libertad contractual ha sido detenida y otra vez el péndulo se está inclinando a su favor.

Además, esta renovada oscilación del péndulo parece establecida para durar por algún tiempo todavía, cualquiera que sea el resultado de las próximas elecciones. Todos los partidos políticos aceptan ahora las virtudes (dentro de límites variables) del libre mercado, del libre contrato, e incluso aquellos que se oponían a las tendencias actuales parecen ahora haber aceptado muchos de los cambios recientes como esencialmente irreversibles.

Estos cambios en el clima político y económico han, por supuesto, producido cambios en el derecho, especialmente en la legislación. Pero el derecho de los contratos es en su mayor parte *common law* y quizás podrá parecer a primera vista que el *common law* continúa en su curso imperturbable y no influenciado por cambios políticos y económicos en el mundo exterior a las cortes. Es poco probable que muchos libros de derecho contractual mencionen el cambio en la dirección política que sobrevino con la elección de 1979. En Inglaterra la tradición de "neutralidad" política de las cortes es tan fuerte que abogados y jueces generalmente no invocan ideologías políticas y económicas al discutir materias de derecho contractual. Pero muchas asuntos legales son profundamente afectados por preconcepciones subyacentes. Es por ejemplo, posible, que haya mucho desacuerdo, incluso respecto del problema básico de si el derecho debería dejar a las partes cometer sus propios errores, o si es una de las funciones del derecho de los contratos proteger a los débiles de su propia necedad. Es por lo tanto necesario, incluso en un texto introductorio como éste, considerar cómo esta nueva inclinación del péndulo en favor de la libertad contractual está siendo justificada, cuales han sido sus implicancias y qué pueden deparar para el futuro.

Parte de la justificación es, por supuesto, puramente política, lo que quiere decir que simplemente ha habido una renovada fe en algunos de los principios subyacentes tras la libertad contractual - mayor creencia en el derecho del individuo a hacer sus propias libres elecciones y la creciente desilusión respecto de la sabiduría de la toma de decisiones colectivas y burocráticas. Así

también, ha habido una tendencia a apartarse del uso del sistema impositivo para fines de redistribución de la riqueza. La benevolencia estatal y el paternalismo son hoy menos populares que lo que eran una o dos décadas atrás. Existe menos confianza en que "el *Whitehall* sabe lo que es mejor", una mayor creencia en que a los individuos se les debería dejar hacer sus propios acuerdos. Todos estos cambios reflejan una renovada fe en la libertad contractual. Al mismo tiempo ha habido un retorno hacia el mayor uso del sistema competitivo - mayores gratificaciones para aquellos que triunfan, y menor ayuda para quienes fracasan. La libertad contractual tiende a producir precisamente estos resultados, desde que la no interferencia en los contratos significa que a aquellos que están mejor situados para beneficiarse del uso de sus habilidades o recursos, se les dejará gozar de sus ganancias. Aún más, estos cambios políticos han conducido a una caecida área de actividad para los principios contractuales a costa de la regulación pública. Como notamos con anterioridad, la declinante importancia asociada a la libre elección en el período entre 1870 y 1980 llevó a muchas actividades a ser manejadas por entes públicos quienes exigieron impuestos para sus propósitos. La reciente privatización de muchas actividades significa que muchos intercambios económicos, los cuales eran hasta ahora materias de derecho público y regulación, estarán ahora una vez más en el área general de los contratos.

Otra justificación para el reciente cambio de dirección - y una que confirma que muchos de estos cambios serán probablemente permanentes - es que el paternalismo es probablemente menos necesario hoy que lo que era cien o incluso cincuenta años atrás. Los británicos son en la actualidad mejor educados y más sofisticados de lo que eran. Están menos necesitados de protección paternalista. Por supuesto siempre existirán unos pocos socialmente inadecuados, flojos e irresponsables, quienes son incapaces de cuidar en forma adecuada de sus intereses económicos; por ejemplo, siempre existirán los desafortunados, los incapacitados y los desempleados; y muchos en estos grupos puede que sigan necesitando alguna protección paternalista, aunque no debería asumirse que aquellos que necesiten apoyo financiero son necesariamente incapaces de manejar sus asuntos una vez que se les ha proveído del apoyo necesario. Pero la gran mayoría de las personas están hoy mucho mejor equipadas para hacer sus propios contrato y para juzgar sus propios intereses, sin protección paternalista. El gran crecimiento de la vivienda y la copropiedad es un factor significativo en esta mayor sofisticación económica. Más aún, la economía es mucho más competitiva y el consumidor tiene muchas más opciones

de las quepunto las tendencias vigentes se deban a estos cambios sociales actuales, y hasta qué punto son el resultado de los valores cambiantes. Más aún, la economía es mucho más competitiva y el consumidor tiene muchas más opciones de las que existían 2 o 3 décadas atrás. Fue sugerido más arriba, que no hace mucho tiempo atrás una persona que quería comprar una casa habría tenido poca libertad de elección con respecto a los términos y tasas de interés de la hipoteca otorgada a la sociedad constructora. Sin embargo, hoy en día una sorprendente variedad de opciones está abierta al consumidor - él puede pedir prestado (por ejemplo) para una hipoteca a una tasa de interés fijo, o a una tasa que es fija por un número variable de años (a su opción) y las tasas varían importantemente de un prestamista a otro.

Aparte de los cambios en los valores políticos ha habido un nuevo gran interés, como ya hemos dicho, en la relación entre las normas legales y sus efectos económicos; y esto ha sido también un importante factor contributivo del actual resurgimiento de la libertad contractual. En forma breve, una vez más se está diciendo que la libertad contractual es un instrumento principal para la eficiencia económica. Aceptar la libre elección de los consumidores significa que los proveedores tienen que producir lo que los consumidores desean y un aspecto esencial de la eficiencia económica es que la sociedad debería producir lo que los consumidores quieren. Subvencionar industrias ineficientes (por ejemplo) significa que son dedicados excesivos recursos a la producción de cosas que los consumidores no desean comprar al verdadero costo social al cual se producen. Así también, leyes como las usadas para restringir la disponibilidad del crédito del consumidor (depósitos mínimos, por ejemplo) y leyes de control cambiario ya no son aceptadas por el gobierno ya que distorsionan el mercado. Han sido por tanto derogadas, y el área de libertad contractual ha sido correspondientemente ampliada.

Para nuestros propósitos, es quizás más relevante notar cómo es hoy reconocido por los economistas que la interferencia legislativa directa en el libre contrato a menudo tiene resultados desastrosamente ineficientes. Por ejemplo, hoy es ampliamente comprendido que muchos años de control de renta y leyes de protección de los arrendatarios han destruido ampliamente el mercado de las propiedades rentadas en Inglaterra. Debido a que los arrendatarios están tan bien protegidos, las rentas proporcionan un muy bajo retorno de ganancia en la inversión del propietario, y consecuentemente pocas personas están preparadas para invertir en la oferta de casas para la renta. Así, esto crea una disminución en el mercado de arriendo de vivienda, el que puede ser bastante

artificial, y el que no existiría si los propietarios fueran autorizados para cobrar una renta más conveniente. Un caso similar es aquel de la legislación de salario mínimo. Aunque el movimiento sindical y el Partido Laborista permanezcan aferrados a la creencia en semejante legislación como una vía de proteger a los sujetos de bajos ingresos, la mayoría de los economistas creen que este tipo de legislación simplemente conduce a la ineficiencia: eleva el costo de los bienes o servicios producidos por las personas de más bajos ingresos por sobre su valor económico adecuado.

Más aún, esa interferencia legislativa en la libertad contractual no sólo conduce a una ineficiencia económica, hoy también es ampliamente pensado que a largo plazo tiende a dañar a las mismas personas a las cuales se propone proteger. La legislación de salario mínimo, por ejemplo, aumenta el costo de los servicios de los empleados para el empleador. Pero incluso el más elemental conocimiento sobre economía demuestra que incrementar los costos es reducir la demanda. Incrementar el costo de la fuerza laboral (esto es los salarios), no es diferente de incrementar el costo de cualquier otra cosa. Si es más costoso emplear trabajadores porque se incrementan los salarios (ya sea como resultado de una ley de sueldo mínimo o por cualquier otra razón) la demanda por lo que los trabajadores producen bajarán, y de este modo la demanda por los propios trabajadores. Entonces un salario mínimo impuesto legalmente puede, a corto plazo, dar al empleado más dinero, pero puede que a largo plazo, lo lleve a perder su trabajo, y en términos agregados, por lo tanto, a un mayor desempleo. Algunos años atrás se le dio amplia publicidad al caso de un propietario de un pequeño café el cual estaba empleando a una mesera a quien le pagaba menos que el salario mínimo prescrito. Cuando los hechos salieron a la luz, a él se le pidió que subiera el salario de la mesera al mínimo. El protestó que a ese valor el café no podía operar con ganancias y entonces fue cerrado. La mesera perdió su trabajo a pesar de que ella estaba dispuesta a trabajar a un menor sueldo. Acá parece claro que ambas partes estarían mejor si se les hubiera dejado hacer su propio contrato sin interferencia legal. También lo estaría el público, porque habría tenido un café el que al menos habría sido frecuentado por una cantidad suficiente de personas como para pagar un salario suficiente para persuadir a la mesera de trabajar allí. Este fue, por supuesto, un caso muy simple, y el efecto económico de tales leyes es a menudo más complejo y difícil de estimar. Pero estas implicancias económicas de la interferencia legislativa en la libertad contractual están indudablemente ayudando a impulsar la nueva oscilación del péndulo de vuelta a los principios de la libertad contractual.

Entonces nuevamente los economistas están tratando de recordar a los abogados (aunque acá, hasta ahora, con poco éxito) que interferir en algunos de los términos de un contrato probablemente sólo afecta el costo de la negociación, y puede que sea inútil, o absolutamente perjudicial para quienes se pretende ayudar - a menos que también los precios estén controlados. Por ejemplo si la legislación "otorga" a los consumidores que compran bienes, derechos establecidos por ley, que los fabricantes no pueden violar, el casi seguro resultado será el de incrementar el precio de los bienes. Muchos consumidores puede que no objeten esta alza de precios, porque son afortunados en comprar los nuevos derechos al mismo precio - pero el argumento económico es que *estos* consumidores no serán afectados por la legislación ya que de todos modos ellos deberían haber sido capaces de comprar esos derechos en la forma ordinaria, si están dispuestos a pagar por ellos. Los consumidores que *serán* afectados por el cambio legal son aquellos que no piensan que los derechos valgan el precio extra, ya sea porque no pueden afrontar el alza, o por cualquier otro motivo. Estos consumidores serán perjudicados por la legislación ya que son privados de la posibilidad de comprar los bienes sin estos derechos asociados. Es (bajo esta perspectiva) como si una ley fuera aprobada forzando a todos los autos a tener ventanas eléctricas, en vez de permitirles que sean vendidas como un opcional extra. Ahora este argumento particular no ha tenido, hasta el momento, un impacto notorio sobre el derecho, o incluso en la forma en que los abogados (y los consumidores) piensan. Y bien puede ser que el argumento sea defectuoso, o no es, por alguna razón, aplicable en ciertos contextos. Pero es un argumento que probablemente sea más oído en el futuro, y ciertamente uno que debe ser enfrentado.

En forma similar, una vez más se está diciendo, como se dijo durante el apogeo de la libertad contractual, que la desigualdad de poder negociador, y cualquier injusticia resultante en los términos contractuales, no es una materia con la cual el derecho contractual debería involucrarse. La desigualdad de poder negociador es una peligrosa justificación para entrometerse en los contratos. Si existe un mercado verdaderamente competitivo, funcionando bien, entonces ahí no debería existir problema alguno relativo a la desigualdad de poder negociador, porque los consumidores siempre pueden ir a otro lugar. Sin duda, que el comprador común no puede realmente tratar de negociar con J. Sainsbury plc. pero si no le gustan las mercaderías o precios allí ofrecidos, puede encontrar muchas otras tiendas vendiendo lo que él desea. En ese sentido incluso el comprador más pequeño tiene tanto poder negociador como el negocio más

grande, mientras exista un verdadero mercado competitivo. La mayor amenaza negociadora de todas, es hacer el negocio en algún otro lugar. Por supuesto, si existe un problema serio de monopolio, y si esa es la fuente de la desigualdad de poder negociador, entonces diferentes medidas pueden necesitarse para introducir una mayor competencia al mercado, o quizás en algunos casos (como con las Empresas de servicio público) algún grado de control público sea necesario. Y esto es, por cierto, concedido incluso por el actual gobierno, el cual al privatizar los suministros de gas y teléfono, los ha sometido a alguna medida de control público, y no los ha dejado totalmente libres para imponer sus contratos al público. En cada caso, por lo tanto, el argumento vale, la desigualdad en el poder negociador es insuficiente fundamento para la interferencia en contratos negociados libremente. Por cierto, el funcionamiento eficiente del mercado *requiere* realmente que a aquellos con gran poder negociador debería serles permitido usarlo - todos aquellos con talentos extraordinarios - sean jugadores de fútbol o cantantes populares, deberían tener derecho a cobrar lo que el mercado está dispuesto a pagar. Cualquiera que tenga algo que otra persona desee muchísimo tiene un poder negociador desigual, si esa cosa posee escasa oferta, y debería tener derecho a venderlo a cuanto pueda obtener.

Esto significa, entre otras cosas, que más o menos en los últimos 15 años el gobierno actual y la política legislativa han buscado, al menos a largo plazo, restaurar la libertad contractual en ciertas áreas en las cuales ha estado mucho tiempo ausente, así como, por ejemplo, control de alquileres y legislación de salario mínimo. Se ha entendido, por supuesto, que los arrendatarios pobres o los trabajadores de bajos ingresos, pueden necesitar apoyo financiero por parte del Estado; pero la tendencia moderna es suministrar ese apoyo directamente a quienes lo necesitan, más que tratando de usar el derecho contractual como un medio redistribuidor de riqueza. Existe claramente mucho acierto en este cambio de política. Si los arrendatarios son tan pobres como para pagar una renta apropiada, entonces necesitan ayuda financiera, pero no es tan claro porqué esta ayuda debería ser suministrada por los propietarios más que por los contribuyentes; y de todos modos, como hemos visto, tratar de forzar a los propietarios a hacer el trabajo de los servicios de seguridad social significa simplemente que habrán muy pocos propietarios, y muy poca inversión en la construcción de propiedades para la renta. De forma similar, si un trabajo es de valor económico tan pequeño que no se le puede pagar un salario que se considere socialmente adecuado, puede haber un caso que requiera ayuda

financiera, pero una vez más, no es claro porqué esta ayuda debería ser otorgada por el empleador antes que por el Estado.

Los modernos escritores de estos asuntos económicos no se han olvidado bajo ningún respecto de algunos de estos argumentos esbozados con anterioridad, que llevaron a la declinación de la creencia en la libertad contractual. El difundido uso de contratos tipo, por ejemplo, y el hecho de que las partes contratantes a menudo no entienden, o ni siquiera leen los términos por los cuales se obligan, es un problema que aún debe ser resuelto, si es que los principios de la libertad contractual van a gobernar una vez más. Un posible acercamiento a estas dificultades es tratar de distinguir entre dos cuestiones por separado. El primer asunto es si una parte contratante realmente ha acordado o aceptado los términos en cuestión. Y el segundo asunto es si aquellos términos son justos o razonables. La primera cuestión se dice que tiene relación con la "justicia procedimental" - se refiere al proceso por el cual los contratos son hechos, es decir, a las reglas que operan dentro del mercado. Es aceptado por todos que este es un asunto propio del derecho de los contratos. Pero la segunda cuestión se refiere a una "justicia sustantiva", y eso se relaciona con el *resultado* de los contratos. Muchos escritores modernos de asuntos económicos insisten en que la segunda cuestión no es una materia apropiada para ser resuelta por el derecho: ellos insisten que la justicia de los términos en sí, siempre se reflejará en el precio, y es de todas formas algo que las partes contratantes deben responder por sí mismas. En cierto sentido, la aceptación de esta distinción entre justicia procedimental y sustantiva es el asunto más fundamental en el derecho contractual moderno y en la teoría, pero como en el caso de muchos asuntos de ideológicos similares, una adherencia rígida a la distinción es difícil en la práctica. Retornaremos a este importante asunto en el Capítulo 16.

El actual estado de los principios y políticas subyacentes a gran parte del derecho contractual es muy difícil de resumir, porque tenemos, en efecto, dos grandes tendencias que corren en direcciones opuestas. Por una parte, tenemos la nueva tendencia hacia un resurgimiento de los principios de la libertad contractual, aunque esto significa un retroceso a las ideas de comienzos del siglo diecinueve. Y por la otra parte, tenemos la antigua tendencia, que deriva de finales del siglo diecinueve, alejada de la libertad contractual. Esta antigua tendencia ha ido quedando sin fuerza, pero de ninguna manera se ha detenido completamente. Por ejemplo, la Directiva de la Unión Europea de Protección al Consumidor, de la que se trata en el Capítulo 16, muestra como en algunas áreas, las ideas y valores contrarias a la libertad contractual se mantienen fuertes. En

todo caso, las antiguas tendencias han dejado como secuela muchas leyes que permanecerán vigentes hasta, y si es que, se disponga de ellas. De este modo aún tenemos grandes cuerpos de derecho derivados del período 1870 -1980, cuando la libertad contractual estaba en decadencia. Tenemos mucha legislación proteccionista y paternalista en las leyes. La mayoría de esta legislación probablemente sería defendida por abogados educados en la atmósfera reinante antes de 1980, aunque parte de ésta ciertamente parece contraria a la tendencia moderna. Gran cantidad de legislación moderna relativa al crédito del consumidor, por ejemplo, es muy proteccionista y paternalista. Interfiere en la libertad contractual a una gran escala. Inevitablemente eleva el costo del crédito del consumidor, porque las compañías financieras deben decidir el costo adicional impuesto por los consumidores que incumplen sus obligaciones contractuales amparados en la legislación. Más aún, existe un riesgo adicional de que semejante legislación realmente proteja al incompetente y al irresponsable a costa del honesto y ahorrativo. Este tipo de "otorgamiento de subvenciones" es una característica común de gran parte de la legislación de este período. Es un producto natural de la filosofía colectivista más que de una filosofía individualista. Parece contrario a la moderna oscilación que retorna a la libertad contractual. Pero es muy dudoso que muchos abogados o políticos estén aún preparados para aceptar que esa legislación está obsoleta. La legislación que confiere derechos a las partes contratantes es aún frecuentemente anhelada por muchas organizaciones disputarían el argumento de los economistas que tal legislación va a menudo en contra de los intereses de quienes parecen ser beneficiados por ella.

Entonces una vez más, la decadencia de la importancia del derecho contractual durante el período que va entre 1870 y 1980 fue acompañado, como notamos con anterioridad, por un crecimiento de la importancia de otras áreas del derecho de obligaciones, particularmente el derecho de los daños extracontractuales y el derecho de la restitución, los cuales frecuentemente imponen obligaciones sobre aquellos que no han acordado soportarlas. Parece poco probable que la nueva oscilación del péndulo de la libertad contractual haga que los abogados den vuelta sus espaldas a estos desarrollos legales de los últimos treinta o cuarenta años, los cuales, a la mayoría de los abogados les han parecido innovadores y productores de mejorías. No obstante, existen señales (como veremos más adelante) de un leve retroceso del derecho de daños extracontractuales en la década pasada, de las áreas en donde tal derecho se superpone particularmente al derecho contractual. Por cierto, hasta

ese punto, parece que en la tendencia actual está una vez más reviviendo la primacía del derecho contractual sobre otras áreas del derecho de obligaciones .

Por supuesto el retorno a la libertad contractual no significa que "todas" las leyes y toda la interferencia legislativa en los contratos privados sean ahora sospechosas incluso para aquellos que apoyan enteramente las nuevas tendencias. Parte de esta legislación, como vimos anteriormente, se basa en amplios asuntos de interés público, y no es motivada por ideologías paternalistas y proteccionistas. Entonces, leyes como el *Agricultural Holdings Act*, que son diseñadas para promover la agricultura Británica, aún pueden ser halladas interfiriendo con la libertad contractual, como vimos antes. Nuevamente, no hay un rechazo del moderno entendimiento del problema de las "externalidades" - no hay un retorno a la creencia simplista de los años anteriores a 1870 de que todos los contratos privados deben ser considerados en interés público simplemente por estar en interés de las partes. Por otra parte, es indudablemente cierto que la oscilación moderna ha llevado al abandono de muchas antiguas intervenciones en el libre contrato, que descansaban en vastos asuntos políticos, como las antiguas regulaciones que restringían la amplitud de los contratos de compras a plazo, legislación de control de precios, y controles de cambio de moneda.

Pero otras intervenciones de la libertad contractual se mantienen, y generalmente no se considera que sean incongruentes con las nuevas tendencias. En parte, esto se debe al hecho paradójico que la libertad contractual puede ser utilizada de muchas formas que en realidad limitan la libertad contractual. Así, leyes que parecen restringir la libertad contractual pueden en realidad estar diseñadas para intensificar otras áreas de la libertad contractual. Así, por ejemplo, el *common law* y las normas contenidas en leyes diseñadas para derribar o incluso prohibir prácticas restrictivas de la competencia (como veremos en más detalle en el Capítulo 17) Son en cierto sentido interferencias con la libertad contractual, pero ellas en realidad tienen el propósito de intensificar la libertad contractual a través de resguardar el funcionamiento del mercado.

En cuanto a lo que toca al derecho común, los cambios en el clima político y económico de la última década también parecen estar teniendo cierta influencia, como se verá con más detalle después. Existen indicios, por ejemplo, de que en contratos entre comerciantes, la antiquísima jurisdicción de equidad para aliviar a una parte contratante de los efectos de cláusulas penales severas o de cláusulas relativas a la pérdida de derechos, está siendo cercenada. Si estos cambios de dirección verdaderamente justificables, desde un punto de vista sofisticado de lo que la libertad contractual requiere, es una cuestión que

será tratada más adelante. Así también, hay señas de que las cortes están hoy (como sus predecesores de comienzos del sigio diecinueve) cada vez más reacios a usar el derecho para proteger a las partes contratantes débiles de las consecuencias de su propia necedad o ignorancia. Y quizás también hay indicios de que existen dudas cada vez mayores respecto del uso del derecho contractual como dispositivo de redistribución. Ciertamente existe un notable paralelo entre el régimen de los principios del mercado a nivel político y el régimen de la libertad contractual en la esfera legal. En ambos casos, la fuerte tendencia actual consiste en examinar primero la situación del mercado, y después considerar si existe una justificación adecuada para la interferencia en sus resultados, o no. La evasión de estos dos puntos, antiguamente tan común, ahora parece ser desaprobada, y con cierta justificación, ya que no significó una contribución para la claridad del pensamiento.

6.- LOS PROPOSITOS DEL DERECHO DE CONTRATOS.

Después de esta larga introducción, bastará con discutir brevenmente un tema, que también podría ser tratado en extenso, esto es los propósitos del derecho de contratos. Como se deduce de la introducción anterior es evidente que el derecho de contratos sirve muchos propósitos, y llevaría a confusión el insistir (como algunos juristas y teóricos han hecho) que el derecho de contratos esta diseñado simplemente para facilitar los planes privados, o sostener el funcionamiento del mercado, o hacer cumplir las promesas, o protger las expectativas razonables. Sin duda que realiza estas cosas, las que son un elemento central en el derecho de contratos. El derecho contractual facilita y fomenta el intercambio libre y voluntario, al reconocer que dichos intercambios tienen efecto legal y hacer cumplir los acuerdos para efectuar dichos intercambios. De este modo el derecho de contratos le da efecto legal al funcionamiento del mercado, permitiendo así que las actividades de negocios y comerciales sean realizadas eficientemente.

El derecho de contratos también facilita y fomenta la planificación privada de los individuos al dar efecto a las intenciones con las que ellos hacen sus transacciones y planes. Esto permite a los individuos no sólo operar eficazmente como consumidores en el mercado, sino que también en muchos aspectos privados de sus vidas que pueden tener implicancias económicas, aunque ellos no están dentro el mercado como es normalmente entendido. Al cumplir con sus propósitos fundamentales el derecho de contratos tiene tres aspiraciones principales.

Primero, está inspirado en el deseo de hacer cumplir las promesas y proteger las expectativas razonables que son generadas por las promesas y por otras formas de conducta.

En segundo lugar, el derecho de contratos está fuertemente influenciado por las instituciones subyacentes del derecho de bienes, por lo tanto, mientras reconoce y hace cumplir transacciones cuyo objeto es la transferencia de propiedad, en general no apoya ni reconoce transferencias de propiedad o dinero que una persona ha obtenido sin algún intercambio; las transferencias sin intercambio son ampliamente concebidas como constitutivas de un enriquecimiento injusto de una parte a expensas de la otra, y esto es comúnmente un fundamento para la intervención legal. Usualmente esto es hecho invocando un cuerpo separado de normas conocido como el Derecho Restitutorio, pero el derecho de contratos es en si mismo fuertemente influenciado y afectado por a idea que el enriquecimiento injusto no debe ser permitido. Hay, en efecto, un sentido en el que todso el concepto de enriquecimeinto injusto es la imagen espejada del concepto de intercambio.

En tercer lugar, el derecho de contratos está también diseñado para prevenir cierto tipo de perjuicios, particularmente los de naturaleza económica, o al menos para compensar a aquellos que sufren dichos daños. Sobre todo intenta proteger a aquellos que han confiado razonablemente en las promesas o conducta de otros, del daño o pérdidas en que habrían incurrido como consecuencia de su confianza. En este respecto, el derecho de contratos se superpone con gran parte del derecho de daños extracontractuales el que también trata de protger a aquellos que son dañados por otros. Pero mientras sólo una pequeña parte del derecho de daños extracontractuales dice relación con daños que son consecuencia de haber confiado razonablemente, este es el principal punto de atención del rol del derecho de contratos en la prevención de perjuicios.

Porque el primero de estos tres propósitos es usualmente visto como como la función fundamental e incluso exclusiva del derecho de contratos, hay una tendencia (como hemos visto) a tratar las obligaciones contractulaes como auto impuestas, y por lo tanto a pensar que el rol del derecho de contratos de decidir que promesas deben hacerse cumplir y como hacerlo es relativamente uno menor. Entones, las ideas de la comunidad sobre justicia y equidad son vistas como teniendo un rol pequeño en el derecho de contratos. En esto, hay una base de verdad con respecto al primero de los tres propósitos principales del derecho de contratos, pero incluso en relación a este objetivo, el punto es por lo general

fuertemente exagerado.Mientras que muchas obligaciones contractuales son auto impuestas, otras no lo son; Es, por ejemplo, la ley y no las partes quien debe decidir que tipo de expectativas son razonables y por lo tanto deben ser protegidas. En cualquier evento, la totalidad del sistema de acciones del derecho-decidir que tipo de acciones conceder a la parte por el incumpliemiento del contrato-es mayormente impuesto por la ley, y no decidido por las propias partes.

Con respecto al segundo y tercer propósito del derecho de contratos, es de algún modo evidente que ellos requieren una infusión sustancial de los valores de la comunidad. Decidir que tipo de enriquecimiento es injusto y que tipo de confianza es razonable, elementos esenciales en el segundo y tercero de los propósitos principales del derecho de contratos, son materias que deben ser hechas con referencia a las ideas de justicia y razonabilidad de la comunidad.

(1) "Introducción a la filosofía del derecho" (reimpreso en New Haven, 1961), p. 236

(2) El tema es tratado detenidamente en mi obra "El Apogeo y Decadencia de la Libertad Contractual" (Oxford, 1979).

(3) Para un famoso fallo en este sentido, ver *Pringting and numerical Registering Co. vs. Sampson* (1875) LR 19 Eq a 465.

(4) *Saunders vs. Anglia Building Society* (1971) AC 1004.

(5) Además, como se verá más adelante (ver *post,* p. 196) no son sólo los documentos *"firmados"* los que pueden volverse obligatorios en este sentido. Todo tipo de condiciones contractuales pueden ser, y a menudo son, incorporadas en contratos "por referencia" incluso sin ser vistas por una (o incluso ambas) de las partes.

(6) En *Liverpool CC vs. Irwin* (1977) AC 239 la Cámara de los Lores se enfrentó a un arrendamiento de una autoridad local que fue registrado en un documento que simplemente listaba lo que se le exigía hacer al arrendatario y lo que le estaba prohibido. Nada se decía respecto de las obligaciones del arrendador. A partir del *Housing Act 1980* existieron términos "implícitos" establecidos por ley rigiendo algunas de estas materias.

(7) (1956) AC 696.

(8) *Hirji Mulji vs. Cheong SS Co* (1926) AC 497 a 510.

(9) Ibid., a 509.

(10) Dahl vs. Nelson, Donkin & Co. (1881) 6 App. Cas. 38 a 39.

(11) N. 7 supra, a 728.

(12) ver *Roy vs. Kensigton and Chlesea Family Practitioner Committee* (1992) I AC 624.

(13) Ver Lord Diplock en *The Hannah Blumenthal* (1938) AC 834, 916. Como veremos más adelante (ver post p. 148), muchos casos de este tipo son resueltos bajo una doctrina llamada *"estoppel"* pero cualquiera que sea la calificación atribuida al caso, el resultado es proteger a alguien que ha confiado en otro a pesar de que no existe un contrato completo en el sentido legal.

(14) *The Goring* (1988) 1ALL ER 6411.

(15) Ver *Lipkin Gorman vs. Karpnale* (1991) 2 AC 548; *Woolnich Building Society vs. IRC* (1993) I AC 70.

(16) (1980) AC 37.

FORMA Y SUSTANCIA EN EL DERECHO DE CONTRATOS

Patrick S. Atiyah

1. Dos tipos de razonamiento jurídico. A través de las diversas áreas del derecho parece posible encontrar dos tipos diferentes de razonamiento jurídico. Existen, por una parte, razones de sustancia[1]. Si estamos considerando, por ejemplo, la norma que debería regir una situación determinada, podemos ponderar una variedad de argumentos relacionados con la conveniencia o corrección de esta norma o aquella. Si nos preguntamos si una persona debería pagar daños a otra podemos tomar en cuenta una serie de consideraciones que discurrirán en una dirección o en otra, pero esas serán en su mayoría razones de sustancia que apuntan directamente al fondo de la cuestión. ¿Cometió el demandado, por ejemplo, un delito contra el demandante? ¿Causó su conducta algún daño al demandante? Y así sucesivamente. Igualmente si estamos diseñando un procedimiento para la resolución de controversias, podemos ponderar las razones para permitir que las partes recurran a tales o cuales evidencias, o para que dispongan de tales o cuales oportunidades para apelar, razones de sustancia relacionadas todas con los objetivos que intentamos alcanzar.

Por otra parte, con frecuencia nos encontramos invocando razones de un diverso carácter, a las que llamaré razones de forma. En el caso paradigmático, una razón de forma se traduce en la exigencia de escrituración, sello, o tal vez de inscripción o atestado de algún tipo. Cuando un determinado acto no cumple con ciertas formas estipuladas, el derecho priva de eficacia a dicho acto. Un testamento, por ejemplo, o un contrato que necesita constar por escrito, pueden declararse nulos o inejecutables si las formalidades del caso no han sido observadas. En tales casos no nos detenemos por regla general (si bien puede haber excepciones) a preguntarnos si la falta de cumplimiento del requisito formal resulta compensada por alguna razón sustantiva que sugiera reconocerle fuerza jurídica al testamento o contrato. Una vez que la regla jurídica de ineficacia por falta de forma queda claramente establecida, la aplicación de dicha regla deja fuera de consideración los argumentos sustantivos a favor de la validez o ejecución del acto. Las razones de tipo formal son comúnmente asociadas a los documentos escritos. Pero existen muchos otros ejemplos de razones del mismo tipo aun cuando normalmente no nos referiríamos a ellas como razones formales. Por ejemplo, un juez que sigue un precedente vinculante puede hacerlo sin tomar en cuenta los argumentos sustantivos a favor del principio en virtud del cual dicho precedente fue establecido. Si el precedente es realmente vinculante para él, y si él realmente acepta el principio de *stare decisis*, el juez no se detendrá ni por un instante a considerar qué razones sustantivas podrían esgrimirse a favor de la decisión contraria. En lo que a él concierne, la pregunta no cabe plantearla. Carece de sentido sopesar un conjunto de factores con el otro, poner en un lado de la balanza la autoridad del precedente y del otro el peso del razonamiento que pueda hallarse en el

---

Traducido de: P. S. Atiyah. *Essays on Contract*. Oxford: Clarendon Press, 1986, Chap. 5, pp. 93-120. Traducción: Rodrigo Soto Silva.

1 Veasé R. S. Summers, *Two Types of Substantive Reasons: The Core of a Theory of Common Law Justification*, 63 Cornell Law Rev. 707 (1978). Las distinciones que estoy haciendo difieren de las establecidas por el Profesor Summers en ese trabajo; sin embargo, en otras partes el Profesor Summers distingue entre dos concepciones del derecho como "Reglas" y "Razón". Veasé, *Working Conceptions of the Law, 1 Law And Philosophy* 267 (1982). Esta distinción es más cercana a la que yo utilizo en el presente texto. Desde que éste texto fue escrito, el Profesor Summers y yo hemos estado trabajando en una investigación (Tentativamente titulada *Forma y Sustancia en el Derecho Anglo-Americano*) que desarrolla el presente tema en varias dimensiones nuevas.

precedente particular. La razón de forma que el precedente lo obliga simplemente excluye la consideración de cualquier razón contraria. O nuevamente, considérese el caso en que una demanda es desestimada bajo el estatuto de limitaciones (*limitations*). También en ese caso, si el estatuto es adecuadamente invocado y es efectivamente aplicable, éste deja fuera de discusión todos los demás factores. Al menos a primera vista, la corte no sopesa la fuerza del caso promovido por el demandante con la fuerza del alegato de limitaciones. Si el alegato es bueno, todo otro argumento es irrelevante.

La naturaleza de las razones de este tipo ha sido discutida e iluminada en una importante serie de trabajos por Joseph Raz, pero el profesor Raz las denomina "razones de segundo orden" o *pre-emptive* o *protected reasons* más que razones de forma como yo pretendo denominarlas. Nada desde luego depende de meras cuestiones terminológicas, pero es comprensible por qué el profesor Raz elude el término que yo he decidido usar. El razonamiento formal tiene hoy mala prensa entre abogados; en realidad, a menudo se lo trata como la misma cosa que el razonamiento formalista como si ambos fueran ejemplos de mal razonamiento en un rango que va de lo malo a lo muy malo. No niego que existen muchos ejemplos, tanto en el derecho como en lo demás, de razonamiento formalista, el cual es en verdad deplorable. Pero en mi opinión deberíamos distinguir entre razonamiento formal y razonamiento formalista. El razonamiento formal no es *per se* injustificado o erróneo, mientras que el razonamiento formalista sí lo es. Ahí donde las razones de sustancia deberían ser consideradas por quien debe tomar la decisión, y éste se rehúsa a considerarlas, cualquier razón formal que él pueda dar en apoyo de su decisión estará fuera de lugar y será injustificable, y de ahí que se la pueda calificar con justicia de formalista. Un juez, por ejemplo, que afirme que está obligado a tomar una cierta decisión debido a la existencia de un precedente, incluso cuando el precedente es claramente distinguible de acuerdo con las reglas y prácticas aceptadas que gobiernan esos asuntos, está dando una razón formal que no sólo es mala, sino fuera de lugar. Él está siendo formalista. Del mismo modo, un juez que aplica automáticamente una regla sin siquiera detenerse a considerar si es posible o justificable crear una excepción a la regla está razonando de un modo formalista. Nadie puede negar que los libros de derecho están llenos de ejemplos de razonamiento formalista de este tipo. Pero el hecho de que las razones formalistas estén siempre equivocadas no nos autoriza a saltar a la conclusión de que todo razonamiento formal es malo. En realidad, parece bastante claro que el derecho utiliza, y por mi parte argumentaré que utiliza correctamente, razones de forma en todo tipo de situaciones.

Una vez que observamos al derecho y al razonamiento jurídico en detalle, encontramos una gran cantidad de ejemplos de razones de forma que no pueden ser condenadas como malas o formalistas. Leyes, precedentes, sentencias, veredictos del jurado, y decisiones administrativas adoptadas por cuerpos legales, todos ellos plantean cuestiones de este tipo; así ocurre también con el matrimonio y los contratos si bien en formas algo diferentes. Si bien el caso del contrato es aquel al que prestaré mayor atención, quisiera poner la discusión en contexto examinando primero algunos de estos otros ejemplos.

Permítanme considerar primero algunos ejemplos verdaderamente simples relacionados con las fuentes del derecho. El ejemplo más simple de una razón formal invocada en apoyo de una decisión es que así lo exija una regla legal clara. Cuando ese es el caso, el juez

que aplica la ley no tiene motivo para considerar la sustancia o mérito de los argumentos. Dentro de ciertos límites y volveré sobre la cuestión de los límites en un momento es claro que la aplicación de una simple regla legal sólo excluye la consideración de argumentos que serían relevantes en una indagación de puro *common law*, al menos cuando no exista autoridad vinculante. Carecería de sentido plantear argumentos que sugieran que la ley conduce a resultados injustos, a resultados extraños, o anómalos. Asimismo, dentro de ciertos límites carece de sentido y es improcedente sugerir que la aplicación de la ley podría no producir los resultados queridos por el legislador. Es aún más improcedente argumentar que, cualquiera haya sido la posición predominante cuando la ley fue aprobada, las circunstancias han cambiado, o incluso la composición del Parlamento se ha modificado, con lo que aplicar la ley iría en contra de la moral o los puntos de vista actuales, o sería contrario a la opinión de la mayoría existente en la Asamblea legislativa actual. En todos los casos comunes la ley es, y se entiende que sea, una fuente de derecho de carácter formal en el sentido de que ella excluye la consideración de argumentos contrarios a los resultados de su aplicación. La ley no es sólo una razón adicional a ser tomada en cuenta por el juez, que podría inclinar la balanza en alguna dirección, o ser contrapesada por argumentos contrarios. La ley excluye los argumentos contrarios.

Por supuesto, como ya he dicho, todo esto es el caso sólo dentro de ciertos límites. Si una ley es ambigua, o poco clara, o si produce consecuencias que resultan groseramente anómalas o absurdas o tal vez gravemente injustas, entonces los tribunales pueden abstenerse de aplicar esa ley. Los límites dentro de los cuales un juez puede hacer eso, varían según el país, la época, e incluso qué duda cabe de juez en juez. Tales límites han sido discutidos por jueces y autores. No pretendo ahondar en esa discusión. Para mis propósitos es suficiente tener presente que existen ciertos límites, y que todo el mundo reconoce su existencia. Si una ley pudiera ser dejada de lado por el juez cada vez que éste sienta que ella conduce a un resultado que no se daría de no mediar dicha ley, entonces la ley dejaría de ser una fuente del derecho de carácter formal en el sentido en que vengo utilizando el término. Las razones que subyacen a la ley serían simples argumentos adicionales a ser considerados conjuntamente con las demás razones de sustancia relacionadas con el caso. Pero el hecho de que aquellas razones hayan sido plasmadas en una ley no les daría ninguna primacía sobre otras razones.

Algo muy similar ocurre tratándose del sistema de precedentes. De acuerdo con la teoría del *common law*, algunos precedentes son vinculantes para ciertos tribunales en el sentido de que el juez que conozca del siguiente caso está obligado a seguir el principio del primer caso sin examinar las razones de sustancia a favor o en contra de dicho principio. En este caso, sin embargo, tiende a darse un mayor escepticismo del que existe en el caso de la ley. Como todos sabemos, no todo lo dicho en un caso es considerado parte de la *ratio decidendi* que es vinculante para los jueces subsiguientes; más aún, el siguiente tribunal podrá distinguir el caso anterior; y en cualquier caso, existen a menudo reglas y prácticas en relación con la jerarquía relativa de los tribunales que permiten a un tribunal declarar ciertos precedentes obligatorios sobre otros. De nuevo, la práctica en estas cuestiones difiere dependiendo del país, la época, e incluso el específico juez; y por mi parte no subestimaría la trascendencia de estas variaciones en la práctica. Más aún, pienso que es probablemente correcto afirmar que, especialmente en ciertos ámbitos del derecho, un precedente es a menudo tratado como una razón adicional para decidir en una cierta forma, más

que como una clase diferente de razón que excluye la consideración de todas las demás razones. Con todo, al final del día, dondequiera que el *commomn law* opera, pienso que debe haber algunos casos en los que el tribunal se siente obligado a seguir el precedente establecido por un tribunal superior, única y exclusivamente porque ese precedente es considerado una fuente formal para la decisión, que deja fuera de consideración toda razón de sustancia. Por cierto, si olvidamos el efecto del precedente en los tribunales inferiores y pasamos a considerar su efecto en otros operadores jurídicos, o incluso en los ciudadanos, resulta todavía más claro que los precedentes son considerados habitualmente como una fuente jurídica de carácter formal que obliga a todos, por lo menos hasta que otras cortes se pronuncien sobre el mismo asunto. Una vez más, no me olvido de los límites. Se entiende que los precedentes deben ser utilizados en forma inteligente. El abogado que asesora a su cliente, incluso el ciudadano común que lee un caso si es que eso ocurre pueden decir: 'claramente este precedente no se aplicará a esta situación particular; el resultado es demasiado absurdo, y los jueces nunca tuvieron en mente ese tipo de caso'. Aun así, no importa cuáles sean los límites dentro de los que opera la doctrina del precedente, ésta claramente conduce en algunos casos a una razón de forma para la decisión de casos posteriores.

En el caso de los precedentes, muchos abogados se han rebelado ante las implicancias de la doctrina. Bentham, por ejemplo, condenó el sistema de precedentes como uno que conduce a la 'actuación sin razón, a la franca exclusión de la razón y por consiguiente a una oposición a la razón'[2]. Y en tiempos más recientes, un sinnúmero de abogados han protestado particularmente tal vez en los Estados Unidos frente al esclavizante sistema de precedentes como una abdicación de la razón. En el caso de la ley, los argumentos tienden a adoptar una forma diferente. Dado que la relación entre el poder legislativo y la judicatura es normalmente un dato constitucional o un punto de partida, hay pocos juristas que protesten frente a la abdicación de la razón que se exige de parte del juez cuando éste debe aplicar una ley, en todo caso, quizá debería añadir, en Inglaterra. Pero hay quienes, desde luego, argumentarían que por lo menos en lo que respecta a la aplicación de la ley ésta tiene por objeto dar efecto a las intenciones de la legislatura, y que por consiguiente una ley nunca debería interpretarse en una forma abiertamente contraria a la intención del legislador. Si bien es cierto ello ocurre en ocasiones. ¿Debería condenarse esto, en la misma forma en que condenaba Bentham el respeto a los precedentes como 'actuar… contra la razón'? A primera vista podría en efecto merecer la misma condena. Si un amigo me escribe una carta pidiéndome que haga algo por él, pero, interpretando la carta al modo de un Parke B., ejecuto algo un tanto distinto *aun cuando mi leal propósito es llevar a cabo su intención*, mi actuación podría resultar indefendible, en verdad incoherente.

2. Razonamiento formal: el ejemplo del matrimonio. Luego de estos ejemplos de razonamiento formal tomados del proceso judicial paso a considerar el mismo fenómeno en el ámbito del derecho privado. No obstante, de nuevo postergaré el tema del contrato pues primero quiero considerar un área en la que se dan cuestiones muy similares concretamente, el derecho matrimonial. La analogía con el matrimonio puede servirnos para ilustrar y hacer más

[2] Véase *Collected Works of Jeremy Bentham*, Constitutional Code (F. Rosen y J.H. Burns ed., 1983), vol. i, 434.

plausibles algunas de las sugerencias que luego haré en relación con el contrato.

El matrimonio es, por cierto, un estatus jurídico que en los sistemas jurídicos más sofisticados sólo puede crearse siguiendo determinados procedimientos formales. Pero no quiero centrarme en las formalidades requeridas para crear el estatus conyugal sino más bien en el sentido en que el matrimonio en sí mismo constituye una razón formal para la adopción de múltiples decisiones. Para entender este punto basta preguntarse qué ocurriría si la institución legal del matrimonio fuera abolida una posibilidad que ha sido explorada en algunos escritos[3]. Es evidente que ello no evitaría y probablemente ni siquiera desalentaría a hombres y mujeres de vivir juntos tal como lo hacen ahora, en una variedad de formas, y por diversos períodos de tiempo. En algunos casos la cohabitación se extendería sin duda por toda la vida; en otros, ésta sería muy breve. Pero ¿por qué, se preguntará, necesitamos la formalidad de tratar a algunas parejas de convivientes en la forma especial que el matrimonio implica? Después de todo, no es posible seguir respondiendo que el propósito del matrimonio consiste en hacerle saber al mundo que los contrayentes han asumido el compromiso de una unión para toda la vida, cuando sabemos y ellos saben que la misma ley les proporciona vías de escape bastante expeditas e indoloras. Aquellos que abogan por la total eliminación del estatus marital como una formalidad legal innecesaria no niegan que siempre será necesario tomar decisiones en relación con los derechos de propiedad y obligaciones de manutención en los casos de separación, divorcio o muerte. Sin embargo, observan, dado que tales decisiones deberán tomarse, especialmente en los casos de separación y divorcio, tomando en cuenta todas las circunstancias del caso duración del matrimonio, ingresos y comportamiento de las partes, número y edades de los hijos si los hay, etcétera realmente hace muy poca diferencia que el estatus formal de casado sea reconocido o no. El argumento, como se advierte, conduce a afirmar que el matrimonio no es más una buena razón formal para determinar las obligaciones de los convivientes, ya sea en atención a la continuidad de la convivencia o a lo que suceda una vez que ésta termine. Estas cuestiones, se asume, sólo pueden zanjarse considerando razones de sustancia, y no razones puramente formales (y por tanto, dentro de ciertos límites, concluyentes).

Desafiante y poco familiar como puede parecerle este argumento al abogado típicamente conservador, el mismo posee cierta plausibilidad, especialmente en relación a la propiedad y a cuestiones de manutención como las que surgen con la separación y el divorcio, cuando es hoy día una práctica común, en Inglaterra al menos, tratar las respuestas a esas cuestiones como dependientes mucho más de la ponderación general de razones de sustancia. Pero existen, a mi modo de ver, muchas otras cuestiones jurídicas, concernientes tanto al derecho público como privado, cuya resolución no sería tan sencilla de eliminarse el estatus formal de casado. ¿Qué ocurre, por ejemplo con los derechos a accionar bajo la Ley de Accidentes Fatales u otra legislación similar?

No hay problema, responden los partidarios de esta propuesta. Podemos simplemente basar estos derechos en la dependencia de *facto*, y no en el vínculo legal. Bien, eso podría funcionar en las demandas de dependencia, pero desde luego muchos países contemplan otros tipos de derechos legales a demandar en caso de muerte de un esposo, donde se contempla el pago de alguna compensación por el daño moral ocasionado por la perdida en sí

[3] Véase E. M. Clive, *Marriage: An unnecessary Legal Concept?* en *Marriage and Cohabitation in Contemporary Society*, ed. Eeklaar and Katz (1980).

misma. No es tan fácil ver cómo podría pagarse tal compensación cuando no exista una razón de forma como el matrimonio que la justifique. Por lo menos se podría entender necesario indagar si la cohabitación fue exitosa o no en orden a otorgar una indemnización por daño moral ante la perdida del conviviente. Pero dejaremos eso pasar para preguntarnos ahora que ocurriría con la sucesión intestada. Aquí nuevamente, se dice, el derecho podría funcionar sobre la base de la convivencia más que del estatus formal del matrimonio. ¿Cómo funcionaría? Supongo que podríamos fijar un límite en el cual, digamos, veinte años de convivencia conferirían al conviviente sobreviviente máximos derechos en la sucesión intestada, y los beneficios se irían reduciendo proporcionalmente para períodos más breves de convivencia. Pero eso no parece del todo correcto. Supóngase que uno de los convivientes muere después de un brevísimo período de convivencia, su pareja recibiría entonces una muy pequeña sucesión intestada. Seguramente, se argumentará, las *intenciones* de las partes deberían contar para algo. La convivencia casual, realizada sin el propósito de una relación permanente o de largo tiempo, ¿no debería distinguirse de la convivencia seria, planeada y entendida como un compromiso de por vida? En realidad, si el matrimonio fuera abolido parece que tendríamos que reconocer los derechos de los convivientes a expresar sus intenciones en alguna otra forma que el derecho pudiera tomar en cuenta. Sin embargo, observan los partidarios, tales métodos ya existen. Un conviviente sólo tiene que otorgar un testamento para dar efectos a sus intenciones. El problema con eso, desde luego, es que no toma en cuenta el simple hecho de que contraer matrimonio le permite a uno evitarse el tener que otorgar testamento si uno no quiere hacerlo, o lo pospone hasta que sea ya demasiado tarde, como hace mucha gente, sin que de ello se sigan consecuencias totalmente inaceptables en la sucesión intestada.

Pero entonces, se podría objetar, ¿qué ocurre con las relaciones de los convivientes con terceros, con el Estado, por ejemplo, o con fondos de pensiones de empleadores? Las leyes de seguridad social normalmente quieren saber si el titular de los beneficios tiene un esposo, si bien es verdad que un dependiente de *facto* puede a veces tener derechos similares. Los fondos de pensiones podrían, supongo, protegerse mediante declaraciones de voluntad adecuadas. Pero entonces surgen todavía otros problemas, incluido el de tratar de definir ¿qué es convivencia?[4] ¿Qué hay del caso de un hombre y una mujer que viven en diferentes casas pero pasan los fines de semana juntos? Supóngase que hacen eso por diez años. Supóngase que viven bajo el mismo techo pero no tienen una vida sexual normal ¿Cómo podría decidir el derecho si tales personas deben o no ser tratadas como convivientes para algún fin determinado? Una respuesta, por cierto, consiste en facultarlos para producir alguna declaración por virtud de la cual ellos debieran ser tratados como convivientes para todos los efectos jurídicos. Pero en ese caso lo que hemos hecho es, en efecto, reinstaurar el estatus matrimonial y sería más simple no abolirlo desde el principio.

El propósito de esta discusión ahora debería estar claro. El matrimonio es seguramente un excelente ejemplo de una razón de forma para la toma de múltiples decisiones. Son tantas y tan diversas las cuestiones que surgen en relación a como debemos tratar a quienes son parte de un tipo de relación que resulta

---

4 Véase S. M. Cretney, *The Law Relating to Unmarried Partners From the Perspective of A Law Reform Agency*, in Eekelaar and Katz (ed)., op.cit., ch. 36.

extremadamente conveniente y efectivo desde el punto de vista de los costos hacer que las respuestas dependan uniformemente de una sola y simple proposición formal. ¿Están casados o no? El matrimonio se constituye, así, dentro de ciertos límites, en una razón formal concluyente, por ejemplo a la hora de definir los beneficios de la seguridad social o los derechos de la sucesión intestada. Podríamos, desde luego, abolir el matrimonio como institución legal, y responder a todas estas cuestiones mirando más allá del vínculo formal en busca de razones sustantivas. En verdad, bajo ciertas circunstancias es eso lo que hacemos; en particular tendemos a hacer eso en lo que respecta a la propiedad y derechos de manutención en caso de divorcio. Pero hacer lo mismo en todas las demás circunstancias en que queramos disponer de reglas especiales para convivencias de largo tiempo y también para convivientes cuya *intención* sea la cohabitación duradera sería una opción inmensamente costosa y problemática. Posee en consecuencia el razonamiento formal grandes ventajas.

3. *Razonamiento formal: el caso del contrato*. Es tiempo de dirigir nuestra atención al caso del contrato, y su relación con la distinción entre razonamiento formal y razonamiento sustantivo. Podemos empezar considerando el lugar de las razones de forma en el caso relativamente evidente en el que un contrato requiere para existir, o para probar su existencia, constar por escrito. Piénsese, por ejemplo, en aquellos contratos que son declarados ineficaces por el Estatuto de Fraudes toda vez que éstos no consten por escrito. ¿Cuál es la función de la escrituración en ese caso? Debemos observar, en primer lugar, que en ese caso la falta de escrituración es una razón formal y concluyente para decidir en contra (más que a favor) de la creación o existencia de una obligación jurídica. Tal como sucede con otros casos de razones formales, en principio no ponderamos las razones a favor de reconocer la obligación, poniéndolas en la balanza contra la falta de escrituración. Si el contrato no consta por escrito, la posibilidad de reconocer una obligación jurídica es simplemente dejada de lado. Ello simplemente no es concebible. Pero en principio no hay diferencia entre una razón formal para excluir una cierta forma de responsabilidad, de otra que permita reconocer tal responsabilidad. De modo que cabría esperar que las razones a favor de la adopción de semejante razón formal sean similares a aquellas que ya hemos identificado, tales como la reducción de costos, y la minimización del riesgo de error. En el caso particular del Estatuto de Fraudes puede haber base para dudar si las razones en cuestión son todavía hoy adecuadas excepto en el caso de los requisitos relacionados con la transferencia de la tierra. Esas razones el ahorro de costos y la minimización del error bien pueden haber sido más fuertes cuando el Estatuto de Fraudes fue aprobado en 1677 que las que existen hoy día. Pero incluso en 1825 protestaba Bentham por los efectos del estatuto. Si, dijo, hay bases para sospechar que el contrato que se invoca es fraudulento, o insincero, o fabricado, entonces el estatuto es innecesario. Un contrato fabricado obviamente debería ser declarado fabricado y no originar derecho legal alguno. Si, por otro lado, no existen tales bases para sospechar, prosigue Bentham, simplemente es injusto tratar al contrato como ineficaz[5].

Ese se convertiría en el argumento común contra el Estatuto de Fraudes y era el argumento predominante en Inglaterra en 1954. Pero el argumento es erróneo. No toma en cuenta uno de los principales propósitos de las razones de forma en el derecho. Es bastante fácil manejar casos donde los hechos son bien conocidos, o

[5] Bentham, Treatise on Judicial Evidence (1825), p. 121; 'The Rationale of Judicial Evidence? In Works (Bowring edn, vol. vi, 1843), ch. XXV.

indubitablemente reveladores de que el contrato ha sido hecho o no. Pero una de las principales funciones de las razones de forma es que permiten ahorrar tiempo y problemas, y reducir el riesgo de error que conlleva descubrir la verdad. De manera que el requerir la escrituración de ciertos contratos nos libera de investigar si un determinado contrato verbal fue realmente celebrado o no; y dado que nos ahorra ese problema, nos pone a salvo del error de considerar existente un contrato que en verdad nunca se celebró. Por cierto, la razón formal sólo proporciona ese beneficio al precio de excluir la posibilidad de un contrato verbal en casos en que realmente un contrato de ese tipo se haya celebrado. Por consiguiente, tenemos que tener claro cuál de esos dos errores resulta más probable. El hecho de que en Inglaterra rechazáramos la mayor parte de las disposiciones relevantes del Estatuto de Fraudes en 1954 por las razones equivocadas, no significa que el resultado no fuese el correcto desde un punto de vista político.

Permítanme considerar otro aspecto del razonamiento formal en el derecho de contratos. A partir de la decisión en *L'Estrange v. Graucob*, se ha aceptado que, dentro de ciertos límites, la firma en un documento contractual es una razón concluyente para someter a la parte que ha firmado a los términos del documento. Sujeta a la prueba de fraude, falsificación o *non est factum*, la firma es concluyente. Hace unos diez años, el Sr. J. R. Spencer, en un interesante artículo publicado en *Cambridge Law Journal*, presentó esta conclusión como en conflicto con la teoría fundamental del derecho de contratos. Los contratos, dijo, se suponen basados en el acuerdo; pero la decisión en *L'Estrange v. Graucob* puede obligar a una persona que no ha consentido (ya sea porque no ha leído o no ha entendido) en el documento que ha firmado. El Sr. Spencer ciertamente ha puesto el dedo sobre una importante e interesante pregunta. ¿Por qué razón tratamos a las firmas como concluyentes? Seguramente ello se debe a que una firma es una razón formal del tipo de las que venimos discutiendo. Una firma es, y así lo reconoce el común de la gente, un elemento formal, y su utilidad se vería seriamente reducida si fuera posible no considerarla como una razón concluyente a favor de la responsabilidad contractual por lo menos en circunstancias normales. Como en los demás casos, no nos detenemos a examinar las razones subyacentes de sustancia, ni tampoco confrontamos las razones sustantivas con el hecho de que el documento ha sido firmado. La firma es, siempre dentro de ciertos límites, una razón concluyente, para tratar a aquella parte que ha firmado como obligada por el documento. Sin embargo, lo que es, pienso, menos claro es cuál sea la razón de sustancia que subyace en este caso. La razón habitual para justificar el carácter concluyente de la firma es que éste debe aceptarse para mostrar que la parte firmante ha consentido en el contenido del documento; pero otra explicación posible es que la otra parte debe ser considerada como alguien que ha confiado en la firma. De modo que puede ser erróneo preguntar, como alguna vez preguntara H.L.A. Hart, si la firma es una mera prueba concluyente con respecto al acuerdo, o si ésta es en sí misma un criterio de acuerdo. Pero esto es algo que podemos dejar de lado por cuanto no afecta a la distinción entre razones de forma y razones de sustancia.

Había otro problema que preocupó al Sr. Spencer. Las reglas objetivas de interpretación pueden llevar a los tribunales a construir un contrato en una forma que no se compadece con la actual comprensión o intenciones de las partes. Esto perturba al Sr. Spencer. 'Puede ser aceptable', nos dice, 'para el derecho aplicar a una de las partes un acuerdo en el que ella no consintió; pero seguramente hay algo mal en una teoría que pretenda forzar a ambas partes a cumplir un acuerdo que ninguna de ellas consintió'.

Por mi parte no veo nada muy malo en ese resultado en tanto recordemos que el derecho no está tanto obligando a las partes a entrar en una determinada relación aunque algunas veces pueda llegar a hacer incluso eso como determinando las obligaciones que para cada uno se siguen de aquella relación en la que decidieron entrar. En realidad, si las partes difieren con respecto a cuáles sean sus obligaciones, me parece a priori mucho más probable que la perspectiva justa y razonable sea una que se sitúe entre las pretensiones de ambas partes, que el que una u otra de las partes esté totalmente en lo correcto. Así, resulta perfectamente natural, un lugar común en verdad, para el tribunal sostener que el contrato impone obligaciones a las partes que no equivalen exactamente a lo que cada una pretende.

El punto fue descrito con su habitual claridad y vigor por Holmes:

> Nada [dice] es más cierto que las partes pueden obligarse a través de un contrato a cosas que ninguna de las dos pretendió, y cuando uno no conoce el parecer del otro. Supóngase un contrato debidamente perfeccionado y escrito para dictar una conferencia, en el que no se menciona la fecha. Una de las partes piensa que la promesa será construida para significar, dentro de una semana. La otra piensa que significa, cuando esté listo. La corte dice que significa, dentro de un tiempo razonable. Las partes quedan obligadas por el contrato tal como éste es interpretado por la corte, si bien ninguna de las partes quiso decir lo que la corte declaró que habían dicho.

El próximo punto a tener en cuenta también fue sugerido por Holmes, si bien en este caso con menos claridad de la que era habitual en él:

> Pienso [escribió en una carta de 1896] que en la teoría ilustrada, para la cual ahora estamos listos, todos los contratos son formales (…) No quiero decir sólo que la *consideration* de un contrato simple es tanto una forma como lo es un sello, sino que en la naturaleza de un sistema de derecho razonable (que trata principalmente con *externals*) la elaboración de un contrato debe ser una cuestión de forma, aun cuando los detalles de nuestra legislación debieran cambiarse[6].

Confieso que no estoy muy seguro de lo que Holmes quiso decir al afirmar que 'la elaboración de un contrato debe ser una cuestión de forma', pero hay una manera en la que estas observaciones pueden entenderse. Tal como, en el caso del matrimonio, donde distinguí entre las formalidades usadas para crear el vínculo conyugal, y el estatus conyugal en sí mismo, el cual también constituye una razón para adoptar varias decisiones, asimismo podemos distinguir entre las formalidades (si las hay) usadas para la celebración de un contrato y el contrato en sí mismo como una fuente formal de obligaciones. Es formal en el sentido de que la sola formulación de una promesa, o incluso un acuerdo, el mero acto de alcanzar un acuerdo, el mero acto de expresar una intención, difícilmente pareciera alcanzar el nivel de una razón de sustancia para hacer algo. He argumentado en otro lugar[7] que la obligación contractual rara vez descansa en la sola intención de las partes. En la gran mayoría de los casos, he sugerido, la parte obligada por un contrato ha recibido algún beneficio de la otra parte, o ha inducido a la otra parte a confiar en él ocasionándole perjuicios. Pero la obtención de beneficios de otro no es la única fuente de responsabilidad contractual; ésta puede dar origen a responsabilidad incluso en ausencia de acuerdo o promesa. El derecho de restituciones, o enriquecimiento injusto, tiene que ver con estos beneficios no pactados. El derecho de contratos, supuestamente, se relaciona con beneficios pactados; pero he sugerido que no es obvio que tengamos que trazar una línea entre el contrato y el

---

6 Carta a E. A. Harriman of Northwestern University Law School, citada en M. De W. Howe, Justice Oliver Wendell Holmes (1963), vol. ii, 233.

7 *Contracts, promises and the law of obligations*. En la misma compilación.

enriquecimiento injusto en la forma en que lo hace la teoría tradicional.

Asimismo la responsabilidad por inducción de expectativas no se limita al área del contrato. En daños, y con la ayuda del *estoppel* en sus diversas formas, las expectativas razonables son en sí mismas un fundamento justificado para demandar en derecho, incluso cuando no ha mediado promesa[8]. De modo que en la mayoría de los casos la responsabilidad contractual nace de algo en parte al menos que está detrás del acuerdo o la intención de las partes. Tal como en el caso del matrimonio a veces es necesario ver cómo se han comportado las partes, cuánto ha durado la relación, y así, más que limitar la atención únicamente al estatus formal, del mismo modo en el caso del contrato es a veces relevante examinar por qué fue hecha la promesa, o qué es lo que ha sucedido con posterioridad a la perfección del acuerdo. En verdad, podría decirse que son esas las verdaderas razones de sustancia para hacer cumplir la mayor parte de las obligaciones contractuales. Constituye una buena razón para hacer que A pague 100 libras a B el hecho de que A haya recibido de B mercancías aproximadamente por un valor de 100 libras bajo un contrato de venta. El hecho de que A haya prometido o acordado pagar esa suma, 100 libras, es tal vez una razón puramente formal para llegar a esa misma conclusión. Ésta excluye la consideración de las razones de sustancia. No nos detenemos a preguntar, dentro de ciertos límites, si las mercancías tenían realmente un valor equivalente a 100 libras. Igualmente, cuando una persona es inducida a confiar en la promesa de otra, no nos preguntamos si su confianza era razonable o justificada. Esa cuestión es dejada a un lado como irrelevante ante la promesa expresa. Queda determinada en forma concluyente por la promesa es razonable confiar en la promesa.

Dado que estas razones formales o excluyentes sólo operan dentro de ciertos límites, y dado que hoy en día esos límites parecen estrecharse en forma que a menudo los tribunales parecen estar más dispuestos que antes a examinar el valor o los beneficios obtenidos, o la justificación de la confianza en una promesa, he argumentado en otra oportunidad que una promesa o acuerdo opera más bien como una presunción en ocasiones irrebatible, pero que en ocasiones sí puede rebatirse. Que una persona acuerde pagar 100 libras por tal ítem es, según sugerí, buena evidencia de que su valor es en realidad equivalente a 100 libras; y el hecho que una persona haya hecho una promesa a otra es por lo menos evidencia *prima facie* de lo justificable que resultan las acciones ejecutadas por el destinatario de la promesa sobre la base de la promesa que se le hizo. El Profesor Eisenberg ha objetado que mi descripción falla por cuanto de hecho tal evidencia no es tratada como rebatible. Descontadas algunas alegaciones bien establecidas, ninguna evidencia resulta admisible como demostración de que el precio acordado es demasiado alto o demasiado bajo. El precio acordado es tratado como concluyente. E igualmente, como ha argumentado el Profesor Birks, puede decirse que dado que el derecho requiere el pago del precio, ni más ni menos, el remedio que el derecho proporciona debe ser en cierto sentido 'contractual' y no puede ser restitutorio. Por cierto, como cuestión de derecho estricto eso es correcto, si bien no es claro por qué deberíamos dejar de tomar en cuenta las defensas establecidas tales como *fraud* y *misrepresentation* como irrelevantes para el asunto. En cierto sentido, ellas definen precisamente los límites dentro de los cuales el precio acordado

8 Véase *The Hannah Blumenthal* [1983] 1 ALL ER 34, 49; véase también Barry Reiter, en Reiter y Swan, ed., *Studies in Contract Law* (Butterworth, Canada), Study N° 8.

puede considerarse concluyente. Pero al margen de eso, la objeción que se hace a mi forma de plantear el asunto simplemente no entiende la forma en que las razones formales o concluyentes trabajan en el derecho. El hecho de que una presunción sea irrebatible se considera a menudo como demostrativo de que la dicha presunción es realmente una regla de derecho. Pero la razón de ello puede ser totalmente distinta. Si es el caso, como he argumentado, que las obligaciones contractuales derivan en última instancia, por lo menos en los casos normales, de los beneficios obtenidos, o el perjuicio derivado de confiar en lo prometido, entonces todavía es posible explicar por qué tratamos generalmente a las promesas y a los acuerdos como razones formales y concluyentes para el establecimiento y delimitación de tales obligaciones. Tal como ocurre con otras razones de forma, las obvias razones para tratarlas como concluyentes son que éstas ahorran costos de investigaciones más detalladas y minimizan el riesgo de error.

Resulta útil a estas alturas que consideremos un par de ilustraciones simples para mostrar cuán poderosas pueden ser estas razones, y que llamarlas formales no debiera considerarse como denigratorio de su significado. Permítanme considerar primero un ejemplo de la importancia de tratar a las promesas y acuerdos como concluyentes en lo que respecta a las expectativas razonables. Hubo un tiempo en que, en los contratos de edificación, se esperaba del constructor que éste consultara los documentos del contrato, estudiara el sitio, y entonces decidiera cuánto cobraría por el contrato. Los documentos del contrato tradicionalmente establecían en forma expresa que el constructor no tenía derecho a confiarse en ninguna información que pudiera proporcionarle el cliente sino que estaba obligado a hacer su propio examen y estudios, y basarse en éstos más que en la otra parte del contrato. Por mucho tiempo eso ha sido considerado como concluyente respecto de los derechos y deberes de las partes. Sin embargo, a partir de la decisión en el caso *Hedley Byrne*, y a raíz de sugerencias en orden a que una responsabilidad por daños fundada en un razonable deber de cuidado debía existir incluso entre las partes contratantes, se llegó a plantear si el constructor podría resultar impedido de demandar al cliente por negligencia si el cliente había dejado de proporcionar tal o cual información al constructor. Tal alegación implica, desde luego, que el constructor confió razonablemente en el dueño, más que en sus propias indagaciones, por más que el contrato estipulaba claramente que no debía hacerlo. Uno podría haber pensado, entonces, que el contrato debía considerarse una razón concluyente que descartaba la posibilidad de alegar razonable confianza por parte del constructor en tales circunstancias. El contrato, se diría, sencillamente define de antemano qué clase de confianza será tenida por razonable, y no hay lugar para ninguna indagación de hechos en el caso particular. No obstante en 1972 el Tribunal Supremo de Australia sostuvo que no era ese el caso, y que un deber de cuidado existe entre las partes en ese tipo de relación; pero por supuesto la existencia de semejante deber sólo podría demostrarse mediante un examen cuidadoso y detallado de todos los hechos. Con posterioridad a ese caso, otro caso australiano, *Dillingham Construction Pty Ltd v. Downs*, llevó a la práctica esa doctrina en un caso de negligencia. El resultado fue un prolongado juicio en el que todos los hechos debieron investigarse minuciosamente; el conocimiento que el cliente tenía acerca de las condiciones particulares del sitio, si tal conocimiento podría razonablemente haber sido recogido y puesto a disposición del constructor, qué conocimiento tenía el constructor, quién confió razonablemente en quién, etcétera. Al final ninguna responsabilidad fue establecida, pero el hecho de que

la confianza del constructor en el cliente no fuera razonable o justificable (si es que en verdad existió algún tipo de confianza) sólo pudo establecerse a un enorme costo. Si el punto de vista tradicional con respecto al efecto de un contrato en este tipo de relación se hubiera mantenido, todos esos costos se habrían podido ahorrar tratando al contrato como una razón concluyente para excluir la consideración de tales hechos como irrelevante. No digo que el procedimiento seguido en ese caso fuera injustificado, o incluso indeseable. Únicamente estoy indicando lo costoso que resultó. Tratar esas cuestiones como establecidas en forma concluyente por el acuerdo resulta barato y simple. Con todo, en ocasiones es necesario y el precio debe pagarse. La justicia es con frecuencia una comodidad costosa.

Permítanme ahora observar que exactamente lo mismo puede ocurrir en relación con los beneficios. Todo el derecho de restituciones o enriquecimiento injusto da testimonio de lo difícil que resulta establecer, en muchas circunstancias, si una persona se ha beneficiado con las acciones de otra. Ofreceré un solo ejemplo impactante de los costos adicionales que se derivan del hecho de dejar a un lado las disposiciones del contrato. Este caso proviene de un grupo de casos británicos de fines del siglo XIX y comienzos del XX relacionados con la doctrina del *ultra vires*. Después de que finalmente se estableciera que una compañía no podía entrar válidamente en un contrato fuera del ámbito de poderes conferido por el memorando de la asociación, se llevaron ante los tribunales varios casos en los que compañías habían obtenido préstamos de dinero bajo transacciones *ultra vires*. Los tribunales se mostraron naturalmente reacios a permitir que las compañías retuvieran las cantidades así obtenidas sin tener ninguna obligación de restituirlas, de modo que echaron mano de diversos expedientes para evitar ese resultado. Si, sostuvieron, una compañía solicitó un préstamo de dinero a través de un contrato *ultra vires*, pero el dinero fue utilizado para extinguir deudas de la compañía con terceros, entonces el prestamista puede demandar restitución del dinero hasta el monto del beneficio que éste le haya reportado a la compañía que solicitó el préstamo. Dado que el dinero había sido utilizado para extinguir deudas de la compañía, era bastante fácil concluir que sí le había reportado beneficios a la compañía. Por desgracia, se dieron ciertas circunstancias en las que esta conclusión no resultaba tan simple como podría parecer. ¿Cómo podría uno estar seguro de que la compañía realmente se había beneficiado con el dinero recibido a través del préstamo *ultra vires*? Supóngase, por ejemplo, que el dinero había sido pagado a la cuenta de un banco con el propio dinero de la compañía, y parte de él pagado a terceros acreedores, mientras que otra parte fue destinada a propósitos varios. ¿Quién podría decir si la compañía realmente obtuvo un beneficio, o si, a fin de cuentas, desmejoró su situación a resultas de recibir el préstamo *ultra vires*? En términos más generales, se observará que argumentos basados en el beneficio obtenido por una persona a resultas de algo acontecido en el pasado implican una indagación sobre lo que habría sucedido, si la persona no hubiera recibido dicho beneficio; tales indagaciones implican exploraciones hipotéticas en relación con lo que hubiera sucedido si el curso de los eventos hubiera sido distinto de cómo en efecto fue. En forma análoga a las indagaciones que en ocasiones es preciso hacer tratándose de acciones de daño personal en las que resulta necesario probar la existencia de beneficios colaterales, esas indagaciones hipotéticas son extremadamente problemáticas. Considérese cuánto más simples hubieran resultado estos casos si los contratos en cuestión no hubiesen sido *ultra vires*. En ese caso las compañías que solicitaron el dinero habrían sido responsables

de pagar los préstamos sin que se planteara la más mínima cuestión respecto de si se habían beneficiado con los préstamos o no. La ganancia en simplificación y el ahorro de costos habría sido inmenso. Una vez más, debería ser claro que no siempre podemos evitar ese tipo de indagaciones respecto de si una persona se ha beneficiado con las acciones de otra. Todo lo que quiero mostrar es que permitir a las personas que decidan si algo hecho por ellos constituye en verdad un beneficio por el que se les podrá requerir pagar conlleva una gran simplificación y eficiencia de costos en el derecho. De nuevo, entonces, los argumentos a favor de tratar las razones de forma como concluyentes, y no permitir ninguna exploración de las razones de sustancia, son muy fuertes; aun cuando, en alguna ocasión, la justicia exija que las dejemos de lado.

4. LA DECADENCIA DEL RAZONAMIENTO FORMAL.

No sería difícil multiplicar los ejemplos de las grandes ventajas que se siguen del hecho de que el derecho de contratos trate normalmente muchas de las cuestiones que subyacen al contrato como irrelevantes para las obligaciones de las partes. Al mismo tiempo, todo el mundo debe ser consciente de que el poder de las razones de forma en el derecho de contratos y, en verdad, probablemente en todo el derecho, ha ido decayendo en años recientes. Cada vez más a menudo los tribunales parecen dispuestos a explorar la transacción, abrirla, e ir más allá de las razones de forma, y atender a las razones de sustancia para la creación o negación de las obligaciones. En algunos casos, estos desarrollos en el derecho parecieran conducir a un mero estrechamiento de los límites dentro de los cuales las razones de forma son tratadas como concluyentes. Pero en otros casos es difícil evitar la conclusión de que las razones de forma en sí mismas han sido echadas por la borda, y las cuestiones tratadas como si estuvieran completamente en discusión. La virtual desaparición en el moderno derecho británico de la regla de *parol evidence* parece ser un claro ejemplo de la castración de una razón de forma en casos de contratos. Cuando se argumenta que un contrato escrito no es concluyente con respecto a las obligaciones de las partes, el tribunal parece simplemente poner en la balanza los argumentos a favor y en contra de imponer a una de las partes una obligación derivada de términos verbales, más que tratar a la escrituración como concluyente, incluso dentro de los límites más estrechos. De igual forma, la discrecionalidad legal conferida a los tribunales por la Ley de Cláusulas Abusivas (*Unfair Contract Terms*) de 1977 y otros estatutos similares, envuelve la eliminación del efecto formal o concluyente de varias cláusulas contractuales. Tales cláusulas no se transforman en irrelevantes, pero su efectividad dependerá de un balance de argumentos sustantivos a favor y en contra de la cláusula.

Esta tendencia a abandonar el uso de las razones de forma me parece que es evidente en muchas áreas además del derecho de contratos. Ya me he referido al caso de *R. v. Lee*, en el cual la Corte de Apelación admitió la apelación de alegato de culpabilidad, como otro ejemplo notable del abandono de un tipo de razón de forma en el procedimiento penal. Así también el moderno Estatuto de Limitaciones, el cual confiere a la Corte una total discreción para pasar por alto los límites de tiempo en una acción de *personal injury*, es otro ejemplo del abandono de una razón de forma para disponer de un caso. Hoy día, los límites de tiempo y otras razones de sustancia tienen que ser sopesados al decidir si cabe permitir que una acción de ese tipo prosiga aún después que el plazo normal ha expirado. En otros países, y especialmente en Estados Unidos, me parece que las razones de forma han ido perdiendo alcance y fuerza

concluyente por muchos años; por ejemplo, la doctrina del ataque colateral en el procedimiento penal federal americano es un ejemplo extraordinario de la destrucción del peso formal aparejado a la finalidad de las decisiones de Cortes Estatales.

¿Por qué ha ocurrido esto? Para responder esta pregunta debemos, a mi juicio, retroceder un poco, y volver a plantear nuestra indagación acerca de los propósitos y razones que subyacen al uso de las razones de forma. En este ensayo he intentado identificar diversos propósitos y razones. He puesto énfasis en los costos, y en la minimización del riesgo de error como las dos razones más importantes de carácter general. Pero de ningún modo son esas las únicas razones. Así, he sugerido que en algunos casos, varios otros factores apoyan el uso de razones de forma. En el derecho de contratos, especialmente, se da la circunstancia de que las partes contratantes son las personas más indicadas para decidir qué beneficios valen la pena, y las cortes no deberían impedir que las partes tomen sus propias decisiones en tales cuestiones. Esto difiere del argumento según el cual sería muy costoso para los tribunales investigar si una persona ha obtenido beneficio de un contrato, y de ser así, por cuánto. Asimismo otras razones se aplican a otros casos. Por ejemplo, algunas razones de forma se refieren a la persona indicada para tomar la decisión cuestiones de jurisdicción surgen aquí *cuándo* una decisión debería tomarse, y por qué procedimiento debería plantearse una cierta cuestión. Así también surge el valor de la tranquilidad, finalidad y seguridad en los asuntos humanos. La búsqueda de la justicia es importante para todo abogado, como también para el público desde luego; pero incluso esa búsqueda debe cesar en casos particulares cuando hayamos hecho lo mejor que podamos. Muchas de esas razones, por cierto, se solaparán en ciertas circunstancias.

Ahora, un extendido uso de razones de forma en la adopción de decisiones presupone a mi juicio que las razones de sustancia también serán, o han sido, o cuando menos podrían haber sido, más apropiada y satisfactoriamente consideradas en otra oportunidad, o en alguna otra manera o por alguna otra persona. Cuando impedimos una acción dentro del período de limitación, por ejemplo, lo hacemos en la convicción de que la acción podría y debería haber sido ejercida dentro del período indicado, y entonces juzgada de acuerdo a sus méritos. Si tal suposición es infundada, por ejemplo, porque los estatutos de limitación impiden una acción antes que el demandante pudiera estar advertido siquiera de que tenía una causa para accionar, entonces hay algo que está realmente mal. O de nuevo, cuando una acción es desechada por falta de jurisdicción, la suposición normal es que alguien tendrá jurisdicción. Incluso si ningún otro tribunal puede conocer del caso, de modo que el sujeto activo permanece inerme, la denegación automática por falta de jurisdicción presupone que el asunto podría cuando menos recibir la atención de la legislatura. Pero también aquí, si la cuestión no pudiera ser atendida por ninguna otra corte, y existieran razones para pensar que la legislatura nunca trataría el asunto, la razón de forma jurisdiccional resultaría pobre y eventualmente quizá, inaceptable. Sin duda esta es una de las principales razones que llevaron a la Suprema Corte americana a intervenir en los casos de *apportionment* cuando resultaba evidente que la legislatura no corregiría la cuestión por sí misma.

O una vez más, la interpretación literal, esto es formal, de una ley, ignorando las consecuencias y tal vez las intenciones del Parlamento, puede resultar más justificable si es posible asumir, primero, que las leyes son redactadas y aprobadas con cuidado, de modo que la presunción de que el significado literal refleja correctamente la intención del Parlamento es una presunción

fuerte; y en segundo término, si puede asumirse que el Parlamento actúa en forma razonable pronto a remediar cualquier deficiencia en la ley causada por tal decisión. Si la redacción de las leyes fuera hecha siempre con espantosa ineptitud y oscuridad, y si el Parlamento se rehusara a hacer cualquier cosa al respecto, es seguro que los tribunales no intentarían interpretar las leyes en una forma literal y formal. Es, después de todo, sólo desde que se extendió la práctica de redactar y aprobar las leyes en la forma actual, que los métodos actuales de interpretación fueron adoptados. Del mismo modo, el principio de finalidad de las decisiones, al que todavía se adhiere con bello rigor en Inglaterra, sólo se justifica si podemos asumir generalmente que la mayoría de las decisiones son consideradas cuidadosamente, y emitidas con competencia e integridad, y que las oportunidades para apelar están disponibles. Si esas condiciones no se dan si, por ejemplo, las cortes de apelación no pudieran confiar en que los tribunales inferiores observarán el derecho con razonable competencia e integridad y si, por alguna razón existieran obstáculos para apelar contra tales decisiones, es difícil creer que la finalidad de las decisiones que tanto apreciamos hoy día no resultaría socavada.

Adoptar una decisión por referencia a una razón de forma, después de todo, es rehusarse a considerar las razones de sustancia relacionadas con el asunto. Pero si esas cuestiones de sustancia nunca han de ser consideradas en absoluto, todavía más, si nunca ha existido la oportunidad para considerarlas, entonces adoptar la decisión sobre la base de razones puramente formales equivale a cerrar la posibilidad de que esa decisión se base en los factores que son más directamente relevantes para ella. Las instituciones racionales y un sistema jurídico racional, requieren que no hagamos eso si podemos evitarlo. Ahora me parece que una de las razones por las cuales las razones de forma son hoy menos favorecidas en el derecho de contratos arranca de las dudas crecientes con respecto a si las razones de sustancia que más directamente afectan al resultado han sido alguna vez adecuadamente consideradas por alguien. Si las partes de hecho no han anticipado los eventos que han ocurrido, no es satisfactorio sostener que las palabras literales del contrato puedan ajustarse para cubrir esos eventos, dado que eso equivale a emplear razones de forma ahí donde las razones de sustancia no han sido ni serán consideradas por las partes ni por nadie. Probablemente, asimismo, crecientes dudas acerca de qué tanto leen o entienden los consumidores los contratos impresos sean una de las principales razones que subyacen a la introducción de discrecionalidad legal para manejar cláusulas abusivas de exclusión. También aquí, dar efecto a la cláusula equivale a disponer del caso sobre la base de una razón de forma cuando las razones de sustancia tal vez nunca fueron consideradas en lo más mínimo, y por cierto tampoco pueden serlo ahora dado que es demasiado tarde.

Sin duda, en el área del contrato y quizá también en las demás, existen factores de este tipo que ayudan a explicar por qué tantas razones de forma del tipo del que he venido discutiendo han sido recientemente abandonadas o mutiladas en el derecho. Sin condenar necesariamente todas las manifestaciones de esta tendencia, este proceso me parece uno contra el cual deberíamos estar en guardia. El formalismo, por cierto, tiene un mal nombre; y adoptar una decisión por razones de forma exige un grado de autodisciplina; exige que admitamos que alguna otra persona, algún otro momento, algún otro lugar, algún otro procedimiento, puede ser el correcto para esa decisión. Pero eso a menudo es cierto tratándose de decisiones adoptadas en un sistema jurídico en el que de cada uno pueda esperarse que desempeñe su propio papel. Cuando ese sea el caso, no deberíamos sentir temor de reconocer

que decidir un caso por razones de forma es decidirlo por buenas razones.

145
c. 3

PUBLICACIONES DE LA ESCUELA DE CRIMINOLOGIA
DE LA UNIVERSIDAD DE BARCELONA, BAJO LA DIRECCION DEL
Prof. O. Pérez-Vitoria

CRIMEN Y COSTUMBRE EN LA SOCIEDAD SALVAJE

por

BRONISLAW MALINOWSKI

OBRAS PUBLICADAS
Stephan Hurwitz: CRIMINOLOGIA

...IDAD DE CHILE

EDICIONES ARIEL
Barcelona

41223

Título del original inglés
CRIME AND CUSTOM IN SAVAGE SOCIETY
Routledge & Kegan Paul Ltd. Londres

Traducido por
J. y M. T. ALIER

Primera edición inglesa, 1926.
Primera reimpresión, 1932.
Segunda reimpresión, 1940.
Tercera reimpresión 1947.
Cuarta reimpresión, 1949.
Quinta reimpresión, 1951.

© *Ediciones Ariel, S. L. 1956*

Printed in Spain

Talleres Tipográficos de Ariel, S. L. - Aragón, 255 - Barcelona

A Sir Richard Gregory
Doctor en Ciencias
Editor de «Nature»

el ímpetu primero fué tan poco edificante que los más modernos antropólogos ignoran su misma existencia tanto en la teoría como en el trabajo práctico. En el clásico manual *Notes and Queries on Anthropology*, la palabra *ley* no aparece ni en el índice ni en la tabla de materias y las pocas líneas que se le dedican bajo el título de *Government: Politics*, con todo y ser excelentes, no corresponden en modo alguno a la importancia del asunto. En el libro del malogrado Dr. Rivers sobre *Social Organization*, el problema de la ley primitiva se discute sólo de una manera incidental y, como veremos, a través de la breve referencia que de ella hace el autor, más bien se le excluye que se le incluye en la sociología primitiva.

Esta laguna en la antropología moderna es debida, no a desinterés por la legalidad primitiva, sino, por el contrario, a su excesiva exageración. Aunque parezca una paradoja, es sin embargo cierto que la antropología actual descuida la ley primitiva porque tiene una idea exagerada y, voy a decirlo sin ambages, equivocada de su perfección.

PRIMERA PARTE

LA LEY Y EL ORDEN PRIMITIVOS

I

LA SUMISION AUTOMATICA A LAS COSTUMBRES Y EL VERDADERO PROBLEMA

Cuando nos preguntamos por qué ciertas reglas de conducta, por duras, molestas o desagradables que sean, son obedecidas; qué es lo que hace transcurrir tan fácilmente la vida privada, la cooperación económica y los sucesos públicos; en una palabra, en qué consisten la fuerza de la ley y el orden en la sociedad salvaje, la respuesta no es fácil y lo que la Antropología ha podido decirnos dista de ser satisfactorio. Mientras se pudo sostener la teoría de que el "salvaje" es realmente salvaje, de que éste sigue caprichosa y descuidadamente la poca ley que tiene, el problema no existía. Cuando esta cuestión adquirió verdadera actualidad, cuando se hizo patente que lo característico de la vida primitiva es más bien la hipertrofia que no la carencia de reglas y leyes, la opinión científica viró en redondo; al salvaje se le convirtió, no sólo en un modelo de ciudadano cumplidor de la ley, sino que se tomó como axioma que, al someterse a todas las reglas y limitaciones de su tribu, el salvaje no hace más que seguir la tendencia natural de sus propios impulsos; que de esta manera, por así decirlo, se desliza fácilmente por la línea de menor resistencia.

El salvaje — según el veredicto actual de competentes antropólogos — siente una reverencia profunda por la tradición y las costumbres, así como muestra una sumisión automática a sus mandatos. Los obedece "como un esclavo", "ciegamen-

te", "espontáneamente", debido a su "inercia mental" combinada con el miedo a la opinión pública o a un castigo sobrenatural; o también por el "sentimiento prevalente en el grupo cuando no por el instinto de grupo". He aquí lo que dice un libro reciente: "El salvaje está muy lejos de ser la criatura libre y despreocupada que nos pinta la imaginación de Rousseau. Por el contrario, se halla cercado por las costumbres de su pueblo, encadenado por tradiciones inmemoriales, no sólo en sus relaciones sociales, sino también en su religión, su medicina, su industria, su arte: en pocos palabras, en cada aspecto de su vida" (E. Sidney Hartland en *Primitive Law* pág. 138). Podríamos muy bien aceptar todo esto, aunque es incierto que las "cadenas de la tradición" sean idénticas o similares en arte y en relaciones sociales, en industria y en religión. Pero cuando inmediatamente se nos dice que "todas estas leyes son aceptadas por el salvaje como una cosa corriente que a él ni se le ocurre quebrantarlas", entonces nos vemos obligados a protestar. ¿No es acaso contrario a la naturaleza humana el aceptar cualquier represión como si fuese cosa natural y para el hombre civilizado o salvaje cumplir reglamentos desagradables, pesados y crueles, someterse a prohibiciones, etc., sin que se le obligue? ¿Y que se le tenga obligado usando alguna fuerza o motivo al que él no puede resistir?

De este modo se nos asegura de nuevo que los "métodos intuitivos" o "no deliberados", la "sumisión instintiva" y un misterioso "sentimiento de grupo" son la causa de que haya tanto ley como orden, comunismo y promiscuidad sexual todo de una vez. Esto suena exactamente como un paraíso bolchevique, pero es ciertamente equivocado en lo que hace referencia a sociedades melanésicas que conozco por observación propia.

Y, sin embargo, esta sumisión instintiva, esta conformidad automática de cada miembro de la tribu a sus leyes, es el axioma fundamental sobre el que se basa el estudio del orden primitivo y su adherencia a la ley. Otra de las más destacadas autoridades en esta cuestión, el malogrado Dr. Rivers, se refiere en el libro que ya hemos mencionado a "un método no deliberado o intuitivo de regular la vida social" que, según él, "está estrechamente ligado al comunismo primitivo". Y sigue diciéndonos: "En un pueblo como el melanesio hay un sentimiento de grupo tan fuerte que hace innecesaria cualquier organización social concreta para la ejecución de la autoridad, exactamente del mismo modo que hace posible el funcionamiento armonioso de la propiedad comunal y asegura el carácter pacífico de un sistema comunista de relaciones sexuales". (*Social Organization*, pág. 169).

Una idea similar es la que expresa aun un tercer escritor, un sociólogo, que ha contribuído más a nuestra comprensión de la organización de los salvajes desde el punto de vista de la evolución mental y social que quizás cualquier otro antropólogo viviente. El profesor Hobhouse, hablando de tribus con un nivel cultural muy bajo, afirma que "tales sociedades tienen naturalmente sus costumbres que para sus miembros son percibidas como obligatorias, pero si nosotros entendemos por ley un conjunto de reglas que una autoridad competente se encarga de hacer cumplir valiéndose de lazos personales de parentesco y amistad, entonces una institución como ésta no es compatible con su organización social" (*Morals in Evolution*, 1955, pág. 73). Aquí tenemos la cuestión de la frase "percibidas como obligatorias" y nos preguntamos si estas palabras no encubren y ocultan el verdadero problema en vez de resolverlo. ¿Acaso no hay, por lo menos con respecto a ciertas reglas, un mecanismo que obliga, aunque tal vez no esté reforzado por ninguna autoridad central, sino sólo respaldado por verdaderos motivos, intereses y sentimientos complejos? Por un "mero sentimiento" ¿se pueden hacer efectivas las prohibiciones más severas, los deberes más gravosos y las obligaciones más duras e irritantes? Nos gustaría saber más sobre esta inapreciable actitud mental, pero el autor sencillamente lo da por descontado. Es más, la definición mínima de la ley como "un conjunto de reglas que una autoridad independiente de lazos personales se encarga de

hacer cumplir" me parece demasiado estrecha y que no destaca suficientemente los elementos más pertinentes. Entre las muchas normas de conducta en las sociedades salvajes hay ciertas reglas que se consideran como obligaciones ineludibles de un individuo o grupo hacia otro individuo o grupo. El cumplimiento de tales obligaciones se recompensa por regla general de acuerdo con la medida de su perfección mientras que su incumplimiento repercute sobre el moroso. Si nos hacemos fuertes en un punto de vista tan comprensivo como éste e investigamos la naturaleza de las fuerzas que lo hacen obligatorio, podremos llegar a conclusiones mucho más satisfactorias que si tuviésemos que discutir asuntos de autoridad, gobierno y castigo.

Consideremos otra opinión representativa de este asunto, la de una de las más altas autoridades antropológicas de los Estados Unidos, el Dr. Lowie, y veremos que expresa un punto de vista muy similar: "En general, las leyes no escritas de uso consuetudinario acostumbran a ser obedecidas con mucha mayor diligencia que las de nuestros códigos escritos. Mejor dicho, son obedecidas espontáneamente" (1). Comparar la "diligencia" en obedecer la ley de un salvaje de Australia con un neoyorkino o la de un melanesio con un ciudadano inconformista de Glasgow es un procedimiento peligroso y las conclusiones a que conduce deben considerarse como muy "en general" hasta que, claro está, pierden todo su significado. El hecho es que no hay sociedad que pueda trabajar de un modo eficiente sin que sus leyes sean obedecidas "diligentemente" y "espontáneamente". La amenaza de coerción y el miedo al castigo no afectan al hombre medio, tanto "salvaje" como "civilizado", mientras que por otra parte son indispensables para ciertos elementos turbulentos o criminales de una y otra sociedad. Asimismo hay que tener presente que en cada cultura humana hay cierto número de leyes, prohibiciones y obligaciones que pesan mucho sobre cada ciudadano, exigen gran sacrificio personal y sólo son obedecidas por razones morales, sentimentales o prácticas, pero sin "espontaneidad" alguna.

(1) «Primitive Society», Cap. sobre «Justicia», página 381, edición inglesa.

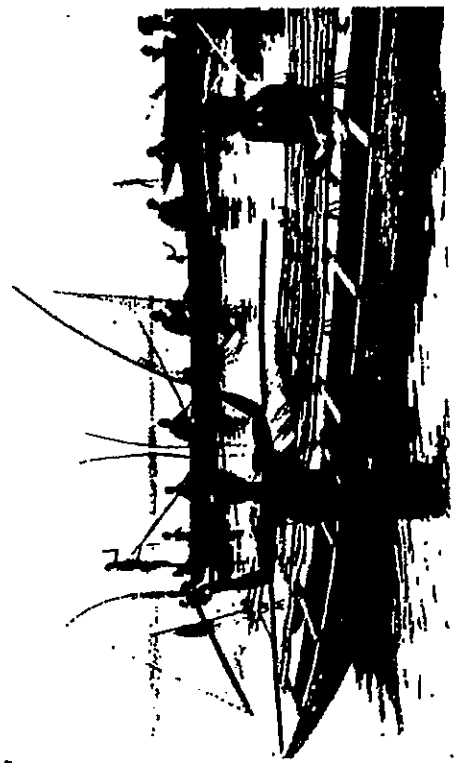
No sería difícil multiplicar los ejemplos y demostrar que el dogma de la sumisión automática a las costumbres de la tribu domina toda la investigación de la ley primitiva. A decir verdad, sin embargo, debe hacerse resaltar el hecho de que cualesquiera imperfecciones de teoría u observación son debidas a las reales dificultades y riesgos que tanto abundan en el estudio de este tema.

La dificultad mayor del problema estriba, según creo, en la misma naturaleza compleja y difusa de las fuerzas que constituyen la ley primitiva. Acostumbrados como estamos a buscar una organización definida de funcionamiento de la ley, administración e imposición del cumplimiento de la ley, también buscamos algo parecido en la comunidad salvaje y, al no encontrar soluciones similares, llegamos a la conclusión de que toda ley es obedecida por esta misteriosa propensión del salvaje a obedecerla.

Al parecer, la antropología se enfrenta aquí con una dificultad similar a la que tuvo que vencer Tylor en su "definición mínima de religión". Al definir las fuerzas de la ley en términos de autoridad central, de códigos, tribunales y alguaciles, llegaríamos a la conclusión de que la ley no necesita que se la haga cumplir en una comunidad primitiva, sino que es seguida de una manera espontánea. Algunos observadores han anotado el hecho de que el salvaje también quebranta la ley algunas veces —pero sólo en ocasiones y raramente— y ello ha sido aprovechado por fundadores de teorías antropológicas que siempre habían sostenido que la ley criminal es la única ley de los salvajes; que cuando el salvaje observa las prescripciones de la ley bajo condiciones normales, cuando ésta es seguida y no desafiada, lo es como máximo, parcial, condicionalmente y aun sujeta a evasiones; que no se le obliga a cumplirla por ningún motivo poderoso como miedo al castigo o sumisión general a toda tradición, sino por

móviles psicológicos y sociales muy complejos; todo esto es un estado de cosas que la antropología moderna hasta ahora ha pasado por alto. En la exposición que sigue trataré de demostrar en un sector etnográfico, el Noroeste de Melanesia, las razones por las que observaciones de naturaleza semejante a las que yo he llevado a cabo debieran extenderse además a otras sociedades con el objeto de adquirir alguna idea sobre sus condiciones legales.

Examinaremos los hechos dentro de un concepto muy amplio y muy flexible del problema que se plantea ante nosotros. En la investigación de derecho y de fuerzas legales, trataremos simplemente de descubrir y analizar todas las reglas concebidas y seguidas como obligaciones efectivas, de descubrir la naturaleza de las fuerzas que limitan y obligan, y de clasificar las reglas de acuerdo con la manera en que son hechas efectivas. Veremos que por un examen inductivo de los hechos, llevado a cabo sin idea preconcebida alguna ni definición apriorística, podremos llegar a una satisfactoria clasificación de las normas y reglas de una comunidad primitiva, a distinguir claramente la ley primitiva de otras formas de costumbre, y a un concepto nuevo, dinámico, de la organización social de los salvajes. Dado que los hechos de derecho primitivo descritos en esta obra han sido registrados en Melanesia, el área clásica de "comunismo" y "promiscuidad", de "sentimiento de grupo", "solidaridad de clan" y de "obediencia espontánea", las conclusiones que podremos establecer —que liquidarán todos estos términos pegadizos y lo que ellos significan— serán de interés especial.



GRABADO I

Canoas de pesca en la laguna.

II

LA ECONOMIA DE LOS MELANESIOS Y LA TEORIA DEL COMUNISMO PRIMITIVO

El Archipiélago Trobriand, que está habitado por la comunidad melanésica a que me refiero, se extiende al noroeste de Nueva Guinea y consiste en un grupo de islas coralinas que rodean una amplia laguna. Las partes llanas están cubiertas de suelo fértil, los peces pululan por la laguna, y tierra y agua ofrecen además fáciles medios de intercomunicación a sus habitantes. Por lo tanto, estas islas mantienen una densa población principalmente dedicada a la agricultura y la pesca, pero también experta en varias artes y oficios, y activa en el comercio y el cambio.

Como la mayoría de los habitantes de las islas de coral, pasan una gran parte de su tiempo en la laguna del centro. En un día de calma aparece llena de vida con canoas llevando gente o productos, o dedicados a uno de los muchos métodos de pesca que les son propios. Un conocimiento superficial de estas actividades podría dar la impresión de arbitrario desorden, anarquía y completa falta de sistema. No obstante, pacientes y cuidadosas observaciones pronto nos revelarían el hecho de que los nativos, no sólo tienen definidos sistemas técnicos de pescar y complicados convenios económicos, sino que además disponen de una estrecha organización en sus equipos de trabajo, así como una división fija de funciones sociales.

Así, pues, veríamos que dentro de cada canoa hay un

hombre que es su verdadero propietario, mientras que el resto actúa como su tripulación. Todos estos hombres, que por regla general pertenecen al mismo sub-clan, están ligados unos a otros y a los individuos de su mismo poblado por obligaciones mutuas; cuando toda la comunidad sale a pescar, el propietario no puede negar su canoa. O bien debe salir él mismo o dejar que vaya alguien en su lugar. La tripulación está asimismo obligada a él. Por razones que pronto se verán claras, cada hombre debe ocupar su sitio y cumplir con la tarea que le corresponde. Del mismo modo cada participante recibe su parte correspondiente de lo que se ha cogido, es decir, equivalente al servicio que ha prestado. Vemos, pues, que la propiedad y uso de la canoa consiste en una serie de obligaciones y deberes concretos que unen a un grupo de gente y lo convierten en un equipo de trabajo.

Lo que hace que las condiciones sean todavía más complejas es que los propietarios y los miembros de la tripulación tienen el derecho de ceder sus privilegios a cualquier pariente o amigo. Esto se hace a menudo pero siempre a cambio de retribución, de retorno. Tal estado de cosas puede aparecer como muy igual al comunismo a cualquier observador que no capte bien todas las complicaciones de cada transacción: parece como si la canoa fuese propiedad de todo un grupo y usada indiscriminadamente por toda la colectividad.

El Dr. Rivers nos dice textualmente que: "uno de los objetos de la cultura melanésica que es usualmente, por no decir siempre, el tema de propiedad común es la canoa" y más lejos, refiriéndose a esta afirmación, habla de "hasta qué gran alcance los sentimientos comunistas concernientes a la propiedad dominan al pueblo de Melanesia". (*Social Organization*, págs. 106 y 107). En otro trabajo, el mismo autor nos habla de "la conducta socialista o incluso comunista de sociedades tales como éstas de Melanesia" (*Psychology and Politics*, páginas 86 y 87).

Nada sería más erróneo que tales generalizaciones. Hay una distinción y definición estricta de los derechos de cada

uno y esto hace que la propiedad lo sea todo menos comunista. En Melanesia tenemos un sistema compuesto y complejo de guardar la propiedad que de ningún modo participa de la naturaleza del "socialismo" o del "comunismo". Así, una compañía por acciones modernas podría ser calificada de "empresa comunista". De hecho, cualquier descripción de una institución salvaje en términos tales como "comunismo", "capitalismo" o "compañía por acciones", tomados de las condiciones económicas actuales o de controversia política, no puede sino inducir a error.

El único procedimiento correcto es el de descubrir el estado de cosas legal en términos de hechos concretos. Así, la propiedad de una canoa de pesca de Trobriand debe ser descrita según la forma como dicho objeto es construído, usado y considerado por el grupo de hombres que lo producen y disfrutan de su posesión. El dueño de la canoa, que al mismo tiempo actúa como jefe del equipo y mago pescador, tiene que pagar ante todo la construcción de una nueva embarcación cuando la vieja ya no sirve, y al mismo tiempo tiene que conservarla en buen estado, aunque en esto le ayude el resto de la tripulación. En esto están bajo obligaciones mutuas de comparecer cada uno a su puesto, mientras que cada canoa debe salir cuando se ha proyectado una pesca colectiva. La embarcación es utilizada de manera que cada asociado tiene derecho a ocupar determinado lugar en ella, lo que implica ciertos deberes, privilegios y beneficios. Cada cual tiene su puesto en la canoa, su tarea asignada y disfruta del correspondiente título, ya sea de "patrón" o "timonel" o "guardián de las redes" o "vigilante de pesca". Su posición y su título vienen determinados por la acción combinada de categoría, edad y habilidad personal. Cada canoa tiene también su lugar en la flota y su parte a representar en las maniobras de pesca conjunta. Así, viéndolo de cerca, descubrimos en esta ocupación un sistema definido de división de funciones y un sistema rígido de obligaciones mutuas en el que se sitúan lado a lado un sentido del deber y el reconocimiento de la necesidad de

cooperación además de la comprensión del interés propio, de los privilegios y de los beneficios. Así, pues, el sentido de propiedad no puede ser descrito con palabras tales como "comunismo" ni "individualismo", como tampoco refiriéndose a sistema de "compañía por acciones" o "empresa personal", sino por los hechos concretos y las condiciones de uso. Es la suma de deberes, privilegios y servicios mutuos lo que liga a los asociados entre sí y al propio objeto.

De modo que, en relación con el primer objeto que atrajo nuestra atención —la canoa nativa— nos encontramos con ley, orden, privilegios definidos y un bien desarrollado sistema de obligaciones.

III

LA EFECTIVIDAD DE LAS OBLIGACIONES ECONOMICAS

Con objeto de adentrarnos más profundamente en la naturaleza de estas obligaciones, sigamos a los pescadores a la playa. Veamos qué sucede con el reparto de la pesca recogida. En la mayoría de los casos sólo una pequeña proporción de ella se queda entre los naturales de aquel poblado. Por regla general encontraremos a cierto número de habitantes de alguna comunidad de tierra adentro que están esperando en la playa. Vemos cómo reciben sartas de pescado de manos de los pescadores y cómo se las llevan a casa, a menudo a muchas millas de distancia, corriendo tanto como pueden para llegar allí mientras el pescado está todavía fresco. Nos hallamos de nuevo ante un sistema de servicios y obligaciones mutuas basado en un convenio ya establecido entre dos poblados distintos. El poblado de tierra adentro suministra hortalizas a los pescadores y la comunidad costera les paga con pescado. Este convenio es primariamente de índole económica. Tiene además un aspecto ceremonial ya que el intercambio ha de efectuarse de acuerdo con un ritual complicado. Asimismo, tiene su lado legal, un sistema de obligaciones mutuas que obliga al pescador a pagar en la misma moneda cuando recibe un obsequio de su compañero de tierra adentro, y viceversa. Ninguno de los dos puede negarse a este compromiso, ninguno de los dos puede escatimar cuando devuelve el obsequio y ninguno de los dos debe retrasarse en hacerlo.

¿Cuál es la fuerza motivadora que respalda estas obligaciones? Los poblados costeros y los de tierra adentro tienen que contar respectivamente el uno con el otro para el suministro de alimentos. Los nativos de la costa nunca tienen suficientes hortalizas mientras que los de tierra adentro están siempre necesitados de pescado. Lo que es más, la costumbre requiere que en la costa todas las grandes exhibiciones ceremoniales y distribuciones de alimentos, que constituyen un aspecto sumamente importante de la vida pública de estos nativos, sean hechas con ciertas variedades de hortalizas especialmente grandes y sabrosas que sólo crecen en las fértiles llanuras del interior. Por otra parte, lo importante de una buena distribución y fiesta es el pescado. Así, a todas las demás razones de peso que valoran los alimentos respectivamente más raros, se añade una dependencia artificial culturalmente creada de un distrito para con el otro. De modo que, en conjunto, cada colectividad necesita mucho de sus asociados. Si previamente, en alguna ocasión, éstos se han mostrado culpables de negligencia, saben que de una forma u otra serán severamente castigados. O sea que cada comunidad tiene un arma para hacer valer sus derechos: la reciprocidad.*

Y ésta no está limitada al intercambio de pescado por hortalizas. Por regla general estas dos colectividades dependen una de la otra también en otras formas de comercio así como en otros servicios mutuos. Por consiguiente, cada cadena de reciprocidad se va haciendo más fuerte al convertirse en parte y conjunto de un sistema completo de prestaciones mutuas.

IV

RECIPROCIDAD Y ORGANIZACION DUAL

Sólo he hallado un escritor que apreciara íntegramente la importancia de la reciprocidad en la organización social primitiva. El destacado antropólogo alemán, Prof. Thurnwald, de Berlín, reconoce claramente "die Symmetrie des Gesellschaftsbaus" (simetría de la estructura social) y la correspondiente "Symmetrie von Handlungen" (simetría de las acciones) (1). A lo largo de su monografía, que es quizás la mejor relación de la organización social de una tribu salvaje existentente, el Prof. Thurnwald nos muestra cómo la simetría de la estructura social y de las acciones llena la vida de los nativos. Sin embargo, su importancia como forma de obligación legal no es expuesta de un modo explícito por el autor, quien parece más consciente de su base psicológica "en sentimiento humano" que de su función social como salvaguarda de la continuidad y adecuación de los servicios mutuos.

Las viejas teorías de dicotomía tribal, las discusiones sobre los "orígenes" de "fratrias" o "mitades" y de la dualidad en subdivisiones tribales, no entraron nunca en los fundamentos interno o diferencial del fenómeno externo de la partición en mitades. La reciente consideración de la "organización dual"

(1) «Die Symmetrie von Handlungen aber nennen wir das Princip der Vergeltung. Dieses liegt tief verwurzelt im menschlichen Empfinden — als adaquate Reaktion — und ihm kam von jeher die grösste Bedeutung im sozialen Leben zu». (*Die Gemeinde der Bánaro*, Stuttgart, 1921, pág. 10).

por el malogrado doctor Rivers y su escuela padece del defecto de buscar causas recónditas en vez de analizar el fenómeno propiamente dicho. El principio dual no es ni el resultado de "fisión", "separación", ni el de cualquier otro cataclismo sociológico. Es el resultado íntegro de la simetría interna de todas las transacciones sociales, de la reciprocidad de servicios, sin los cuales no hay colectividad primitiva que pueda existir. Una organización dual puede aparecer claramente en la división de una tribu en dos "mitades" o ser completamente destruída —pero yo me aventuro a pronosticar que cuando se lleve a cabo una investigación cuidadosa, se encontrará que la simetría de estructura en cada sociedad salvaje es la base indispensable de obligaciones recíprocas.

La manera sociológica cómo se ordenan las relaciones de reciprocidad las hace todavía más estrictas. Los intercambios entre dos comunidades no se llevan a cabo de un modo casual, fortuito, como de dos individuos que comercian al azar el uno con el otro, sino todo lo contrario: cada hombre tiene su compañero permanente en el intercambio y los dos tienen que negociar el uno con el otro. A menudo son parientes políticos o amigos jurados o socios en el importante sistema del intercambio ceremonial llamado *kula*. Además, dentro de cada colectividad los socios individuales son clasificados en sub-clanes totémicos, de modo que el intercambio establece un sistema de lazos sociológicos de naturaleza económica, a menudo combinado con otros lazos entre individuo e individuo, grupos de parentesco y grupos de parentesco, poblado y poblado, distrito y distrito.

Si examinamos las relaciones y transacciones previamente descritas, nos será fácil ver que el mismo principio de mutualismo proporciona la sanción para cada regla. En cada acto hay un dualismo sociológico: dos partes que intercambian servicios y funciones, cada una de las cuales cuidando de que la otra cumpla su parte del compromiso y se conduzca con honradez. El patrón de la canoa, cuyos intereses y ambiciones van ligados a su embarcación, cuida de que reine el orden

en las transacciones internas entre los miembros de la tripulación a la cual representa en sus relaciones exteriores. Para él, cada miembro de la tripulación está bajo obligación en el momento de la construcción e incluso después cuando su cooperación se estima necesaria. Recíprocamente, el dueño tiene que dar a cada hombre el pago ceremonial en la fiesta de construcción; el dueño no puede negar a nadie su lugar en el bote y tiene que cuidarse de que cada hombre reciba su parte correspondiente de la pesca recogida. En esto como en todas las múltiples actividades de orden económico, la conducta social de los nativos está basada en un bien evaluado sistema de toma y daca con cuentas que se llevan mentalmente, pero que siempre se saldarán equitativamente. No hay descuentos en las liquidaciones ni se aceptan tratos de favor; ni hay tal omisión "comunista" de la liquidación de cuentas ni de atribución estricta de participaciones. Todas las transacciones se llevan a cabo de un modo fácil y libre, y las buenas maneras que predominan cubre cualquier inconveniente o disconformidad que pueda presentarse, haciendo que al observador superficial le sea difícil ver el vivo interés propio y el cuidadoso ajuste de cuentas que se efectúa en todo tiempo. Pero para el que conoce a los nativos íntimamente, nada es más claro y patente. El mismo mando que el patrón asume con su canoa es ejercido dentro de la colectividad por el jefe, quien además, por regla general, es el mago hereditario.

V

LEY, INTERES PROPIO Y AMBICION SOCIAL

Casi no es necesario añadir que hay también otras fuerzas poderosas, además de la compulsión de las obligaciones recíprocas, que ligan a los pescadores a sus tareas. La utilidad de esta ocupación, el anhelo que sienten por este alimento fresco y excelente y, por encima de todo, quizá la atracción de lo que para los nativos es un deporte intensamente fascinador, les mueven a seguir pescando más conscientemente y más efectivamente de lo que hemos descrito como una obligación legal. Pero la compulsión social, la consideración por los derechos efectivos y por las pretensiones de los otros ocupan siempre un lugar preeminente en la mente del nativo, así como en sus formas de conducta, una vez esto ha sido bien comprendido. También es indispensable para asegurar el funcionamiento uniforme de sus instituciones, porque a pesar de todo el gusto y atracción que sienten por esto, en cada ocasión hay unos pocos individuos que no se muestran bien dispuestos a colaborar, irritables, malhumorados, obsesionados por algún otro interés — muy a menudo por la intriga — que desearían escabullirse de cumplir con su obligación si ello fuese posible. Cualquiera que sepa lo extremadamente difícil, si no imposible, que es organizar un cuerpo de melanesios, incluso para llevar a cabo una actividad corta y divertida que requiera una acción concertada, y en cambio lo bien y rápidamente que se ponen a la obra para realizar sus empresas habituales, com-

prenderá el papel y la necesidad de la compulsión debida al convencimiento nativo de que otro hombre tiene derecho sobre su trabajo.

Hay todavía otra fuerza que hace los deberes más obligatorios. Ya he mencionado el aspecto ceremonial de las transacciones. Los regalos de alimentos en el sistema de intercambio descrito más arriba deben ser ofrecidos de acuerdo con formalidades estrictas, en medidas de madera especialmente construídas, traídas y presentadas en la forma prescrita, en una procesión ceremonial y con trompeteo de cuernos marinos. Ahora bien, nada tiene mayor influencia sobre la mente de un melanesio que la ambición y la vanidad que van asociadas a la exhibición de alimentos y de riqueza. En la entrega de sus regalos, en la distribución de sus excedentes, experimentan una manifestación de poder y un realce de su personalidad. El nativo de las islas Trobriand guarda sus alimentos en casas mejor hechas y más ornamentadas que las chozas que le sirven de vivienda. La generosidad es para él la virtud más alta, y la riqueza el elemento esencial de la influencia y el rango. La asociación de una transacción semicomercial con ceremonias públicas pautadas suministra otra fuerza obligatoria de cumplimiento a través de un mecanismo psicológico especial: el deseo de exhibición, la ambición de aparecer munificente y la extremada estimación por la riqueza y la acumulación de alimentos.

De este modo hemos podido penetrar algo en la naturaleza de las fuerzas mentales y sociales que convierten ciertas reglas de conducta en leyes obligatorias. Y la fuerza obligatoria no es superflua ni mucho menos, ya que cuando el nativo puede evadirse de sus obligaciones sin pérdida de prestigio o sin posible riesgo de sus ganancias, lo hace, exactamente cómo lo haría cualquier hombre de negocios civilizado. Cuando se estudia más de cerca la "regularidad automática" en el cumplimiento de las obligaciones que tan a menudo se atribuye a los melanesios, se hace patente que hay tropiezos constantes en las transacciones, que hay gran descontento, refunfuñamien-

to y recriminaciones, y que raramente hay un hombre que esté completamente satisfecho con su socio. Pero, en conjunto, el nativo continúa en el consorcio y cada cual trata de cumplir con sus obligaciones, ya que se ve impelido a ello, en parte por inteligente egoísmo, y en parte por obediencia a sus ambiciones y sentimientos sociales. Comparemos al verdadero salvaje, por ejemplo, siempre tan dispuesto a evadirse de sus deberes, fanfarrón y jactancioso cuando los ha cumplido, con el muñeco salvaje del antropólogo que seguiría ciegamente las costumbres y obedecería automáticamente toda regulación. No hay el más remoto parecido entre lo que nos enseña la antropología sobre este tema y la realidad de la vida nativa. Empezamos a ver que el dogma de la obediencia mecánica a la ley impediría al investigador en el terreno constatar los hechos verdaderamente significativos de la organización legal primitiva. Ahora nos damos cuenta de que las reglas de la ley, las reglas que tienen una definida obligación, sobresalen de las meras reglas de la costumbre. También podemos ver que la ley civil que consiste en disposiciones positivas, está mucho más desarrollada que el conjunto de meras prohibiciones, y que el estudio de la ley puramente criminal entre los salvajes pasa por alto los fenómenos más importantes de su vida legal.

Es también obvio que la clase de reglas que hemos considerado, aunque son incuestionables reglas jurídicas obligatorias, no tienen en modo alguno el carácter de preceptos religiosos que siempre son formulados de un modo absoluto y que deben ser obedecidos rígida e integralmente. Las reglas aquí descritas son esencialmente elásticas y adaptables, dejando una laxitud considerable dentro de la que su cumplimiento se considerá satisfactorio. Las sartas de pescado, las medidas de ñame, los manojos de taro, sólo pueden ser evaluados de una manera aproximada y naturalmente las cantidades intercambiadas varían según la abundancia de la pesca o la de la cosecha de hortalizas. Todo esto se tiene siempre en cuenta y sólo la tacañería intencionada, la negligencia o la holgazanería son consideradas como incumplimiento de contrato. Ya que, como hemos dicho, la generosidad es cuestión de honor y de elogio, el nativo corriente hará acopio de todos sus recursos con objeto de mostrarse pródigo en su medida. Sabe además que cualquier exceso de celo y de generosidad será tarde o temprano debidamente recompensado.

Ahora vemos que un concepto estrecho y rígido del problema — una definición de "ley" como la maquinaria de aplicar justicia en casos de transgresión, dejaría fuera todos los fenómenos a que hemos aludido. En todos los hechos descritos, el elemento o aspecto de la ley que es de efectiva compulsión social, consiste en complicados arreglos que hacen que la gente cumpla con sus obligaciones. La más importante de entre ellas es la forma en que muchas transacciones están integradas en cadenas de servicios mutuos, cada uno de los cuales será recompensado en fecha ulterior. La forma pública y ceremonial cómo se llevan a cabo usualmente estas transacciones, combinada con la extremada ambición y vanidad de los melanesios, se suma a las fuerzas que salvaguardan la ley.

VI

LAS REGLAS LEGALES EN LOS ACTOS RELIGIOSOS

Hasta ahora me he referido principalmente a las relaciones económicas, ya que la ley civil trata principalmente de propiedad y riqueza entre los salvajes lo mismo que entre nosotros, pero podremos hallar el aspecto legal en cualquier otro terreno de la vida tribal. Tomemos por ejemplo uno de los actos más característicos de la vida ceremonial — los ritos de duelo por los difuntos —. Al principio, claro está, vemos en ellos su carácter religioso: son actos de piedad hacia el finado causados por el miedo, el amor o la solicitud por el espíritu del muerto. Como manifestación ritual y pública de emoción, forman también parte de la vida ceremonial de la colectividad.

¿Quién, sin embargo, sospecharía un lado legal a tales actividades religiosas? Y, no obstante, en las Islas Trobriand no hay un solo acto funerario, ni una sola ceremonia, que no esté considerado como una obligación del ejecutante hacia algunos de los otros sobrevivientes. La viuda llora y se lamenta en dolor ceremonial, en piedad religiosa y miedo, pero también porque la fuerza de su dolor proporciona una satisfacción directa a los hermanos y parientes maternos del difunto. Según la teoría nativa del parentesco, son los parientes por línea materna los que están realmente afectos. La esposa, aunque vivía con su marido, aunque debe llorar su muerte y aunque a menudo lo hace real y sinceramente, sigue siendo sólo una

extraña de acuerdo con las reglas del parentesco matrilineal. Por lo tanto, su deber hacia los miembros sobrevivientes del clan de su esposo es manifestar, exhibir su dolor de modo aparatoso, guardar un largo período de luto y llevar la quijada o mandíbula de su esposo durante varios años después de su muerte. Esta obligación, sin embargo, no carece de reciprocidad. Cuando tenga efecto la primera distribución ceremonial, unos tres días después de la muerte de su esposo, recibirá el pago de sus lágrimas de manos de los parientes de su marido y será un pago importante; y en las fiestas ceremoniales que se celebran más tarde, también recibirá más pagos por sus subsiguientes servicios de duelo. También debiera tenerse presente que para los nativos el luto es sólo un eslabón en la cadena de reciprocidades de toda la vida entre marido y mujer y entre sus respectivas familias.

VII

LA LEY DE MATRIMONIO

Esto nos lleva al tema del matrimonio, que es extraordinariamente importante para llegar a una verdadera comprensión de la ley nativa. El matrimonio no sólo establece un vínculo entre marido y mujer, sino que también impone una permanente relación de mutualidad entre la familia del esposo y la de la esposa, especialmente el hermano de ella. Una mujer y su hermano están ligados entre sí por lazos de parentesco característicos y muy importantes. En una familia de las Islas Trobriand, una mujer debe estar siempre bajo la tutela especial de un hombre — uno de sus hermanos o, si no tiene ninguno, su pariente materno más próximo —. Ella tiene que obedecerle y cumplir una serie de deberes, mientras él se ocupa de su bienestar y se hace cargo de ella económicamente incluso después de casada.

El hermano pasa a ser el guardián custodio de los hijos de ella que deben considerarle a él y no a su padre como al verdadero cabeza de familia. A su vez, él tiene que ocuparse de ellos y suministrar a la familia de su hermana una considerable proporción de alimentos. Esto resulta tanto más pesado cuanto que el matrimonio, por ser patrilocal, obliga a la muchacha a trasladarse a la comunidad de su esposo, de modo que cada vez que se recoge la cosecha hay un *chassé-croisé* económico general a través de todo el distrito.

Una vez las cosechas recogidas, se procede a la clasifica-

GRABADO II

Nativos del interior tomando sartas de pescado de manos de los pescadores.

ción de los ñames y lo mejor de la cosecha de cada huerto se coloca en una pila de forma cónica. La pila principal de cada huerto es siempre para la familia de la hermana. El único propósito de toda la habilidad y trabajo dedicados a esta exhibición de alimentos es la satisfacción de la ambición del agricultor, ya que toda la colectividad, ¡qué digo!, todo el distrito contemplará los productos cultivados, hará sus comentarios sobre ellos, criticará o elogiará. Según palabras textuales de mi informante, una gran pila quiere decir: "Fijaos en lo que he hecho por mi hermana y su familia. Soy un buen agricultor y mis parientes más próximos, mi hermana y sus hijos, no sufrirán nunca por falta de comida". Transcurridos unos días, se deshace la pila de alimentos, se colocan los ñames dentro de unos cestos y éstos son trasladados al poblado de la hermana, donde se procede a colocarlos exactamente de la misma forma que antes enfrente de la casa-depósito de ñame del marido de la hermana; allí, de nuevo, los miembros de la colectividad verán el montón y lo admirarán. Todo este aspecto ceremonial de la transacción tiene una fuerza obligatoria que ya conocemos. La exhibición, las comparaciones, la evaluación pública, todo esto impone una definida compulsión psicológica sobre el dador, le satisfacen y recompensan cuando el éxito de su labor le permite ofrecer un regalo generoso, pero le castigan y humillan cuando se muestra ineficiente, tacaño o ha tenido mala suerte.

Además de la ambición, la reciprocidad predomina en esta transacción como en todas partes hasta el extremo de que a veces se diría que dar es prácticamente casi tanto como recibir. En primer lugar el marido tiene que recompensar con definidos regalos periódicos cada contribución anual de la cosecha. Más tarde, cuando los niños crezcan, éstos estarán directamente bajo la autoridad de su tío materno; los muchachos tendrán que ayudarle, asistirle en todo, y contribuir con una cuota determinada a todos los pagos que éste tenga que hacer. Las hijas de su hermana hacen poco por él, es decir, directamente, pero indirectamente, en una sociedad matrili-

neal, le proporcionan herederos y descendientes en las dos generaciones siguientes.

Así, pues, si colocamos las ofrendas de la cosecha dentro de su contexto sociológico y con una visión amplia de su relación, vemos que cada transacción está justificada como un eslabón en la cadena de mutuas prestaciones. Y, sin embargo, considerándolo aparte, fuera de su marco, cada transacción nos parece disparatada, intolerablemente pesada y sociológicamente sin sentido y también, sin duda alguna, "comunísticamente". ¿Qué podría ser más económicamente absurdo que esta oblicua distribución de productos agrícolas en que cada hombre trabaja para su hermana y a su vez tiene que depender del hermano de su esposa, y donde se desperdician más tiempo y energía en exhibiciones, alardes y transporte de los graneros que en verdadero trabajo? No obstante, viéndolo de cerca, se comprende que algunas de estas acciones aparentemente innecesarias son poderosos incentivos económicos, que otras suministran la fuerza obligatoria legal, mientras que otras, a su vez, son el resultado directo de las ideas nativas sobre el parestesco. Está claro también que nosotros podemos comprender el aspecto legal de tales cosas sólo si las consideramos integralmente sin exagerar temasiado un eslabón especial de la cadena de deberes recíprocos.

VIII

PREDOMINIO DEL PRINCIPIO DE RECIPROCIDAD

En los capítulos que anteceden hemos visto una serie de ilustraciones de la vida nativa que muestran el aspecto legal de la relación matrimonial, de la cooperación en una partida de pesca, del intercambio de comida entre poblados costeros y de tierra adentro, de ciertos deberes ceremoniales de manifestación de duelo. Estos ejemplos fueron presentados con cierto detalle con el fin de hacer resaltar claramente el funcionamiento concreto de lo que me parece ser el verdadero mecanismo de la ley, de la compulsión social y psicológica, de las verdaderas fuerzas, motivos y razones que hacen que los hombres cumplan con sus obligaciones. Si el espacio lo permitiese, sería fácil presentar estos ejemplos aislados dentro de un cuadro coherente con el objeto de mostrar que en todas las relaciones sociales y en todos los varios dominios de la vida de la tribu se puede descubrir exactamente el mismo mecanismo legal, el cual colóca las *obligaciones forzosas* en una categoría especial que las separa de los otros tipos de reglas consuetudinarias. Sin embargo, en este caso, un rápido y comprensivo examen tendrá que bastarnos.

Tomemos, por ejemplo, en primer lugar, las transacciones económicas: el intercambio de artículos y de servicios se lleva a cabo principalmente dentro de un consorcio establecido o va asociado a lazos sociales definidos o aun acoplado como un mutualismo en asuntos no económicos. Se ve que la mayoría de

los actos económicos, si no todos, pertenecen a alguna cadena de presentes recíprocos y contra-presentes que a la larga equilibran la cuenta y benefician a ambos lados por igual.

Ya he dado cuenta detallada de las condiciones económicas del noroeste de Melanesia en *La Economía Primitiva de los Nativos de las Islas Trobriand* ("Economic Journal", 1921) y en *Argonautas del Pacífico Occidental*, 1923. El capítulo VI de aquel volumen trata de los asuntos que discutimos aquí; por ejemplo, las formas de intercambio económico. Por aquel entonces, mis ideas sobre la ley primitiva no habían madurado suficientemente y presentaba los hechos sin hacer referencia al tema presente, por lo que su testimonio resulta aún más valioso. Sin embargo, cuando describo una categoría de ofrendas como "regalos puros" y bajo este nombre coloco los regalos de marido a mujer y de padre a hijos, está bien claro que cometía una equivocación. De hecho, caigo en el error supuesto más arriba de arrancar el acto de su contexto propio, de no considerar suficientemente la cadena de transacciones. No obstante, en el mismo párrafo suministro una rectificación implícita de mi equivocación al manifestar que "los nativos dicen que un regalo de padre a hijo significa el pago de la relación que el primero sostiene con la madre" (p. 179). Señalo también que los "regalos gratis" a la esposa están basados en la misma idea. Pero la relación verdaderamente correcta de las condiciones — correcta desde el punto de vista tanto legal como económico — sería el abarcar todo el sistema de regalos, deberes y beneficios mutuos intercambiados entre el marido por una parte, y la esposa, hijos y hermano de la esposa por otra. Entonces se vería que de acuerdo con las ideas de los nativos este sistema está basado en un dar y tomar muy complicado y que a la larga los servicios mutuos restablecen el equilibrio (1).

(1) Compárese el justo juicio crítico de mi expresión «regalo puro» con todo lo que implica que hizo M. Marcel Mauss en *L'Année Sociologique*, «Nouvelle Serie», vol. I, págs. 171 y sig. Yo había escrito este párrafo antes de ver las críticas de M. Mauss que concordaban substancialmente con las mías propias. Es satisfactorio para un investigador el ver que ha presentado

La verdadera razón por la que todas estas obligaciones económicas se observan normalmente y se observan además muy escrupulosamente, es que el no cumplirlas coloca a un hombre en una posición intolerable, así como el cumplirlas con retraso o deficientemente le cubren de oprobio. El hombre que persistentemente desobedeciera las reglas de la ley en sus tratos económicos se encontraría bien pronto fuera del orden social y económico — cosa de la que se da perfecta cuenta —. Hoy día se presentan casos en los que algunos nativos, ya sea por holgazanería, excentricidad o espíritu no conformista, habían decidido ignorar estas obligaciones de su estado legal y se ve cómo automáticamente se han convertido en parias y en parásitos de algún blanco.

El ciudadano honorable está bajo la obligación de cumplir con sus deberes, aunque su sumisión no se debe a ningún instinto o impulso intuitivo ni a un misterioso "sentimiento de grupo", sino al detallado y elaborado funcionamiento de un sistema en el cual cada acto tiene su propio lugar y se debe ejecutar sin falta. Aunque ningún nativo, por inteligente que sea, formulara este estado de cosas de una manera general y abstracta ni lo presentara como una teoría sociológica, cada uno de ellos conoce perfectamente su existencia y puede prever las consecuencias en cada caso concreto.

En las ceremonias mágicas y religiosas, casi cada acto — además de su propósito primario — es considerado como una obligación entre grupos e individuos y tarde o temprano tiene que hacerse su pago equivalente o contra-servicio estipulado por la costumbre. La magia en sus formas más importantes es una institución pública en la cual el mago de la colectividad, que por regla general desempeña su cargo por herencia, tiene que prestar sus servicios en favor de todo el grupo. Tal es el caso, por ejemplo, en la magia de cultivo,

tan bien sus observaciones que éstas permiten a otros la refutación de sus conclusiones sacándolas de su mismo material. Y a mí especialmente me ha resultado aun más agradable el constatar que después, al madurar mis ideas, éstas me han conducido independientemente a los mismos resultados que mi distinguido amigo M. Mauss.

pesca, guerra, predicción del tiempo y construcción de canoas. El mago está obligado a ejercer su magia tan pronto como surja la necesidad, en la estación adecuada o en ciertas circunstancias a fin de mantener los *tabús* y a veces incluso dirigir toda la empresa. Por todo esto se le paga con pequeñas ofrendas dadas inmediatamente y a menudo incorporadas en los procedimientos de ritual. Pero la verdadera recompensa está en el prestigio que su posición le confiere (1). En casos de magia menor u ocasional, tales como amuletos para propiciar el amor, ritos curativos, brujería, magia contra el dolor de muelas y de protección de los cerdos, etc., cuando ésta se aplica en favor de otro tiene que estar muy bien pagada y la relación entre cliente y profesional se basa en un contrato dictado por la costumbre. Desde el punto de vista de nuestro argumento actual, debemos registrar el hecho de que todos los actos de magia comunal son obligatorios para el mago y la obligación de ejecutarlos va unida al cargo de mago de la colectividad, que es hereditario en la mayoría de los casos y que siempre es una posición de poder y de privilegio. Un hombre puede declinar su posición y cederla a su sucesor más próximo, pero una vez la ha aceptado tiene que llevar a cabo todo el trabajo inherente a ella, y la colectividad, por su parte, le tiene que dar lo que le es debido.

En cuanto a los actos que usualmente se considerarían como religiosos más bien que mágicos — tales como ceremonias de nacimiento o de matrimonio, ritos de muerte y de duelo, la adoración de los fantasmas, espíritus o personajes míticos —, se ve que también tienen un lado legal claramente ejemplarizado en el caso de los actos mortuorios descritos más arriba. Cada acto importante de naturaleza religiosa es concebido como una obligación moral hacia el objeto que se ve-

(1) Para mayor información referente al estado legal y social del mago hereditario, véase el Capítulo XVII sobre la «Magia» en «Argonautas del Pacífico Occidental», así como las descripciones y referencias diversas sobre la magia de la construcción de canoas, magia de navegar, magia de *kaloma*. Véase también la breve relación de magia de huertos en *Economía Primitiva (Economic Journal*, 1921); de magia de guerra en «Man», 1920 (artículo n. 5); y de magia de pesca en «Man», 1918 (artículo n. 53).

nera: fantasma, espíritu o poder, al propio tiempo que satisface una aspiración emocional del ejecutante; pero, tiene su lugar indiscutible en alguna estructura social y es considerado por una tercera persona o personas como algo que les es debido, vigilado y luego pagado o devuelto por medio de otro servicio parecido. Por ejemplo, cuando con ocasión del retorno anual de los espíritus a su antiguo poblado se hace una ofrenda al espíritu de un pariente muerto, no sólo se satisfacen sus sentimientos y sin lugar a dudas su apetito espiritual que se alimenta de la substancia espiritual de la comida, sino que probablemente se expresa también el propio sentimiento hacia el muerto querido. Además, en ello va envuelta una obligación social: cuando los platos de comida llevan ya algún tiempo expuestos y el espíritu ha terminado con su participación espiritual, el resto — que a pesar de todo no aparece menos adecuado para el consumo ordinario — es ofrecido a un amigo o pariente político viviente quien a su vez devuelve la ofrenda con un regalo parecido algún tiempo más tarde (2). No recuerdo ni un solo acto de carácter religioso que no vaya acompañado de alguna implicación sociológica parecida más o menos directamente asociada a la principal función religiosa del acto. Su importancia estriba en el hecho de que convierte el acto en una obligación social además de constituir un deber religioso.

Podría continuar con el examen de algunas otras fases de la vida tribal y discutir más a fondo el aspecto legal de las relaciones domésticas expuestas más arriba o entrar de lleno en las reciprocidades de las grandes actividades, actos importantes, etc. Pero ahora debe haber quedado ya bien claro que los casos detallados previamente no son casos aislados y excepcionales, sino ejemplos representativos de lo que sucede en cada aspecto de la vida nativa.

(2) Véase la relación que hace el autor de *Milamala*, la fiesta del retorno anual de los espíritus a su antiguo poblado en *Baloma; los espíritus de los muertos en las Islas Trobriand (Journal of the R. Anthrop. Institute*, 1916) — Revista del Instituto Real de Antropología —. Los ofrecimientos de comida que ya he mencionado aparecen en la página 378 de la susodicha revista.

IX

LA RECIPROCIDAD COMO BASE DE LA ESTRUCTURA SOCIAL

Ahora bien, modificando toda nuestra perspectiva y mirando las cosas desde el punto de vista sociológico, es decir, tomando aspectos sucesivos de la constitución de la tribu en vez de examinar las varias clases de sus actividades tribales, sería posible demostrar que toda la estructura de la sociedad de las Trobriand está fundada en el principio de *estado legal.* Con esto quiero decir que los derechos del jefe sobre los individuos particulares, del marido sobre la mujer, del padre sobre el hijo, y viceversa, no se ejercen arbitrariamente ni de un modo unilateral, sino de acuerdo con reglas bien definidas y dispuestas en cadenas de servicios recíprocos bien compensadas.

Incluso el jefe, cuyo cargo es hereditario, basado en tradiciones mitológicas altamente venerables, rodeado como está de un temor semireligioso acentuado por un principesco ceremonial de distancia, humillación y prohibiciones estrictas, que tiene mucho poder, riqueza y medios ejecutivos, tiene que someterse a normas fijas y está ligado por rigurosos vínculos legales. Cuando quiere declarar la guerra, organizar una expedición o celebrar una festividad, debe emitir convocatorias formales, anunciar públicamente su voluntad, deliberar con los notables, recibir el tributo, servicios y asistencia de sus súbditos en forma ceremonial y finalmente retribuirles de

acuerdo con una escala bien definida (1). Bastará mencionar aquí lo que ya se ha dicho del estado sociológico de matrimonio, de las relaciones entre marido y mujer y de la posición entre parientes políticos (2). Toda la división en clanes totémicos, en sub-clanes de carácter local y en comunidades de poblado está caracterizada por un sistema de servicios y deberes recíprocos en el cual los grupos desarrollan un juego de dar y tomar.

Lo que quizás es más notable en el carácter legal de las relaciones sociales es que la reciprocidad, el principio de dar y tomar, también reina supremo dentro del clan, ¡qué digo!, dentro del grupo de parientes más próximos. Tal y como hemos visto, la relación entre el tío materno y sus sobrinos, las relaciones entre hermanos, incluso la relación menos egoísta de todas o sea la de un hombre y su hermana, están todas y cada una de ellas fundadas en la mutualidad y la retribución de los servicios. Es precisamente este grupo el que ha sido acusado de "comunismo primitivo". El clan es a menudo descrito en la jurisprudencia primitiva como la única persona legal, el único cuerpo y entidad. "La unidad no es el individuo, sino el grupo; el individuo no es más que una parte del grupo", según palabras de Mr. Sidney Hartland. Esto es realmente cierto si tomamos en consideración esta parte de la vida social en la que el grupo de parientes — clan totémico, phratria, mitad o clase — participa en el juego de la reciprocidad frente a sus grupos coordinados. ¿Qué hay, pues, de la perfecta unidad dentro del clan? Aquí se nos ofrece la solución universal del

(1) Compárense para mayor detalle los varios aspectos de la jefatura que he señalado en el artículo citado *Economía Primitiva*, op. cit. *(Argonautas)* y los artículos sobre «Guerra» y «Espíritus» también previamente mencionados.

(2) Otra vez debo referirme a mis otras publicaciones en las que he tratado todos estos asuntos con detalle aunque no desde el punto de vista actual. Véanse los tres artículos publicados en *Psyche* de octubre 1923 *(La Psicología del Sexo en las Sociedades Primitivas)*; abril 1924 *(Psico-Análisis y Antropología)*; y enero de 1925 *(Complejo y Mito en el Derecho Materno)*, en los cuales se describen muchos de los aspectos de la psicología sexual y de las ideas y costumbres fundamentales de parentesco y relación. Los dos últimos artículos aparecen uniformados con este trabajo en mi *Sexo y Represión en la Sociedad Salvaje*, 1926.

predominante sentimiento de grupo cuando no del "instinto de grupo" el cual se considera especialmente ufano en esta parte del mundo que ahora nos ocupa habitada por "una gente dominada por un sentimiento de grupo como el que impulsa al melanesio" (Rivers). Como sabemos, éste es un concepto completamente equivocado. Precisamente dentro del grupo de parentesco más próximo es donde florecen las rivalidades, disensiones y el egoísmo más agudo que domina toda la tendencia de las relaciones de parentesco. Pronto volveré a este punto en busca de más hechos, pero de hechos mucho más demostrativos, con objeto de deshacer de una vez este mito de comunismo de parentesco, de la perfecta solidaridad dentro del grupo por descendencia directa, mito que ha sido reavivado recientemente por el doctor Rivers que, por consiguiente, implica cierto peligro de ganar aceptación general.

Una vez mostrado el alcance de los hechos a los que se aplica nuestro argumento y habiendo demostrado bien claramente qué *ley* cubre toda la cultura y constitución de la tribu de estos nativos, formulemos nuestras conclusiones de una manera coherente.

X

DEFINICION Y CLASIFICACION DE LAS REGLAS CONSUETUDINARIAS

Al principio de la Parte Primera se dieron ejemplos de las opiniones corrientes que atribuyen al hombre primitivo una obediencia automática a la ley. A esta suposición van asociadas algunas proposiciones más especiales que son universalmente corrientes en antropología y que, sin embargo, resultan fatales para el estudio del derecho primitivo.

En primer lugar, si el salvaje obedece las leyes de la costumbre debido a su incapacidad de romperlas, entonces no se puede dar una definición de ley, ni se puede señalar ninguna distinción entre reglas legales, morales, maneras y otros usos, ya que la única forma de clasificar las reglas de conducta es en relación con los motivos con que son respaldadas. De modo que basándose en la suposición de que el hombre primitivo presta una obediencia automática a todas las costumbres, la antropología tiene que renunciar a cualquier tentativa de introducir orden y clasificación en los hechos y esto precisamente es la primera tarea de la ciencia.

Ya hemos visto antes que Sidney Hartland considera que las reglas del arte, de la medicina, organización social, industria, etc., están irremisiblemente mezcladas y amontonadas en todas las sociedades salvajes tanto en la propia comprensión del nativo como en la realidad de la vida social. Este punto de vista lo destaca deliberadamente en varias ocasiones: "...La percepción de las semejanzas por parte del salvaje difiere mu-

cho de la nuestra. Él ve un parecido entre objetos que a nuestros ojos no tienen un solo punto en común" (l. c. p. 139). "Para el salvaje... el modo de actuar de una tribu es uno e indivisible... Ellos [los salvajes] no ven nada grotesco ni incongruente en publicar en nombre de Dios un código que combina prescripciones de ritual, moral, agricultura y medicina las cuales nosotros consideramos como prescripciones estrictamente jurídicas... Nosotros, por ejemplo, podemos separar la religión de la magia y la magia de la medicina, pero los miembros de la colectividad no hacen tales distinciones" (pp. 213, 214).

Sidney Hartland, cuando habla de todo esto, da lúcida y moderada expresión a las opiniones corrientes sobre la "mentalidad primitiva prelógica", "las confusas categorías del salvaje", y la general carencia de forma de la cultura primitiva. No obstante, estos puntos de vista sólo cubren una parte del caso, sólo expresan una verdad a medias — en lo que se refiere a la ley, tales conceptos mencionados aquí no son correctos —. Los salvajes tienen una clase de reglas obligatorias no dotadas de ningún carácter místico, que no son anunciadas en el nombre de Dios y que no son respaldadas por ninguna sanción sobrenatural, sino que sólo tienen una fuerza obligatoria puramente social.

Si consideramos la suma total de reglas, convenios y maneras de conductas como el cuerpo de costumbres, no hay duda alguna de que el nativo siente un gran respeto por todas ellas, una fuerte tendencia a hacer lo que los otros hacen, lo que todo el mundo aprueba y, si sus apetitos o intereses no le llaman o impulsan en otra dirección, seguirá la fuerza de la costumbre antes que cualquier otro camino. La fuerza del hábito, el temor de los mandatos tradicionales y su apego sentimental a todo ello, así como el deseo que siente de satisfacer la opinión pública, todo se combina para que la costumbre se obedezca por el solo hecho de serlo. Como puede verse, en esto los "salvajes" no difieren de los miembros de cualquier colectividad completa con un horizonte limitado, tanto si se trata de un *ghetto* (barrio de los judíos) de la Europa Oriental

como de un Colegio de Oxford o de una comunidad fundamentalista del Medio Oeste. Con todo, el amor a la tradición, el conformismo y la fuerza de la costumbre sólo explican hasta un punto muy limitado la obediencia a las reglas entre maestros, salvajes, campesinos o pescadores.

Limitémonos una vez más a los salvajes y veremos que entre los de las Islas Trobriand hay una cantidad de reglas tradicionales para instruir al artesano sobre la forma de ejercer su oficio. La forma inerte como estas reglas son obedecidas sin discusión se debe, por así decirlo, al "conformismo general de los salvajes", pero principalmente son obedecidas porque su utilidad práctica ha sido reconocida por la razón y atestiguada por la experiencia. Además, otras instrucciones sobre la manera de conducirse con los amigos, parientes, superiores, iguales, etc., son obedecidas porque el hombre que se aparta de ellas se siente y aparece ridículo, torpe y socialmente extraño a los ojos de los demás. Éstos son los preceptos de las buenas maneras que están muy desarrollados en Melanesia y a los que todos se adhieren muy estrictamente. Hay asimismo otras reglas para dictaminar las condiciones que deben imperar en los juegos, deportes, diversiones y festividades, reglas que son el alma y substancia de la diversión o actividad y que son observadas porque se siente y se reconoce que cualquier fallo en "seguir el juego" lo echa a perder, lo arruina — es decir, cuando el juego es verdaderamente un juego —. Como se habrá observado, en todo esto no hay fuerzas mentales de inclinación o de interés propio ni incluso de inercia que pudieran contrarrestar algunas de las reglas y hacer que su cumplimiento fuese una pesada carga. Es tan fácil seguir las reglas como no seguirlas y cuando se va a tomar parte en alguna actividad deportiva o de placer, sólo se la puede disfrutar si se obedecen todas sus reglas tanto de arte como de maneras y procedimientos del juego.

También hay normas que pertenecen a cosas sagradas e importantes, las reglas de los ritos mágicos, las pompas funerarias y cosas por el estilo. Estas normas están respaldadas

principalmente por sancionees sobrenaturales y por la fuerte convicción de que con las cosas sagradas nadie debe entremezclarse. Por una fuerza moral igualmente poderosa se mantienen ciertas reglas de conducta personal hacia los parientes, familiares de la casa y otros hacia los que se experimentan fuertes sentimientos de amistad, lealtad o devoción que refuerzan los dictados del código social.

Este breve catálogo no es ninguna tentativa de clasificación. Su propósito principal es indicar claramente que, además de las reglas de la ley, hay varios otros tipos de normas y mandatos tradicionales que están respaldados por motivos o fuerzas, principalmente psicológicas, o en todo caso completamente diferentes de aquellas que son características de la ley en aquella colectividad. Así, aunque en mi examen he enfocado la atención principalmente sobre la maquinaria legal, no he tenido ningún interés en demostrar que todas las reglas sociales son legales, sino todo lo contrario, he querido mostrar que las reglas de la ley no son sino una categoría bien definida dentro del cuerpo de costumbres.

XI

UNA DEFINICION ANTROPOLOGICA DE LA LEY

Las reglas de la ley sobresalen del resto por el hecho de que están consideradas como las obligaciones de una persona y los derechos legales de otra. No están sancionadas por una mera razón psicológica, sino por una definida maquinaria social de poderosa fuerza obligatoria que, como sabemos, está basada en la dependencia mutua y se expresa en un sistema equivalente de servicios recíprocos lo mismo que en la combinación de tales derechos con lazos de relación múltiple. La manera ceremonial como se llevan a efecto la mayoría de las transacciones, que comprende apreciación y crítica públicas, ayuda todavía a su fuerza obligatoria.

Por lo tanto podemos ya descartar la opinión, el parecer de que el "sentimiento de grupo" o la "responsabilidad colectiva" sean la única e incluso la principal fuerza que asegura la adhesión a las costumbres de la tribu y las hace obligatorias o legales. Sin duda alguna, el *esprit de corps, la solidaridad*, el orgullo en la propia comunidad y clan existen entre los melanesios — en realidad sin ellos no hay orden social que se pueda mantener en ninguna cultura alta o baja — sólo quiero prevenir contra conceptos tan exagerados como los de Rivers, Sidney Hartland, Durkheim y otros, que harían de esta desinteresada, impersonal e ilimitada lealtad de grupo la piedra angular de todo el orden social en las culturas primitivas. El salvaje no es ni un "colectivista" extremo ni un "individua-

lista" intransigente sino que es, como todo hombre en general, una mezcla de ambos.

Así, pues, de lo expuesto hasta ahora se deduce que la ley primitiva no consiste exclusivamente, ni tan solo principalmente, en imposiciones, ni toda la ley de los salvajes es ley criminal. Y sin embargo se pretende que con la mera descripción de crimen y castigo el tema del derecho ya está agotado en lo que concierne a la comunidad salvaje. De hecho, el dogma de la obediencia automática, por ejemplo, de la absoluta rigidez de las reglas de la costumbre implica una exageración de la ley criminal en las comunidades primitivas y su correspondiente negativa de la posibilidad de una ley civil. Las reglas absolutamente rígidas no pueden ser aplicadas o adaptadas a la vida ni necesitan ser impuestas, pero pueden quebrantarse. Incluso los que creen en una super-legalidad primitiva deben admitir este punto. De aquí que el crimen sería el único problema legal que se puede estudiar en las comunidades primitivas; no habría ley civil entre los salvajes ni jurisprudencia civil alguna que la antropología pudiese investigar. Este concepto ha predominado en los estudios comparativos de la ley desde Sir Henry Maine hasta las más recientes autoridades en la cuestión, tales como el Prof. Hobhouse, el doctor Lowie y también en Sidney Hartland. De este modo, leemos en el libro de Hartland que en las sociedades primitivas "el núcleo de la legislación es una serie de *tabús*" y que "casi todos los códigos primarios consisten en prohibiciones" *(Ley Primitiva,* p. 214). Y también que "la creencia general en la certeza de un castigo sobrenatural y la alienación de la simpatía del prójimo generan una *atmósfera de terror* que es más que suficiente para prevenir cualquier infracción de las costumbres tribales..." (pág. 8, la bastardilla es mía). No hay tal "atmósfera de terror" excepto quizás en el caso de algunas muy pocas reglas excepcionales y sagradas de ritual y de religión, y por otra parte la infracción de las costumbres de la tribu se previene por medio de un mecanismo especial

Grabado III

Exhibición de dolor obligatoria en los ritos de duelo.

cuyo estudio es el verdadero terreno de la jurisprudencia primitiva.

Hartland no es el único autor que opina así. Steinmetz, en su interesante y competente análisis del castigo primitivo, insiste en la índole criminal de la primitiva jurisprudencia y en la naturaleza mecánica, rígida, no dirigida y no intencional de los castigos impuestos así como de su base religiosa. Sus opiniones son plenamente compartidas por los grandes sociólogos franceses Durkheim y Mauss, quienes añaden una cláusula más: la de que la responsabilidad, la venganza, en realidad todas las reacciones legales están basadas en la psicología del grupo y no del individuo (1). Incluso sociólogos tan agudos y bien informados como el profesor Hobhouse y el doctor Lowie, este último que conoce a los salvajes por experiencia propia, parecen seguir la tendencia general del momento en sus capítulos sobre la justicia en las sociedades primitivas que, por otra parte, son excelentes.

En nuestro propio terreno sólo hemos encontrado hasta ahora mandamientos positivos cuya violación es sancionada, pero no propiamente castigada, y cuya maquinaria ni por los métodos de Procrustes * puede ser retirada más allá de la línea que separa *la ley civil* de *la ley criminal*. Si hemos de etiquetar las reglas descritas en estos artículos de una forma moderna y por lo tanto necesariamente inapropiada, entonces llamémoslas el cuerpo de la "ley civil" de los nativos de las Islas Trobriand.

La ley civil, la ley positiva que gobierna todas las fases

---

(1) Steinmetz, *Ethnologische Studien zur ersten Entwickelung der Strafe*, 1894; Durkheim en *L'Année Sociologique*, I. págs. 353 y sig.; Mauss en *Revue de l'Histoire des Religions*, 1897.

* *N. del T.*: Procrustes, personaje de la leyenda griega, era un bandido del Ática que colocaba a sus víctimas sobre una cama de hierro; si la excedían, cortaba la parte que sobresalía y, si eran más pequeños o cortos, los estiraba hasta hacerles ocupar exactamente la cama. Esta experiencia indica la aplicación no inteligente, a rajatabla, de principios generales sin tener en cuenta las naturales variaciones de casos e individuos.

de la vida de la tribu, consiste por lo tanto en un cuerpo de obligaciones forzosas consideradas como justas por unos y reconocidas como un deber por los otros, cuyo cumplimiento se asegura por un mecanismo específico de reciprocidad y publicidad inherentes en la estructura de la sociedad. Estas reglas de la ley civil son elásticas y poseen una cierta amplitud. No sólo castigan el fracaso sino que premian a los que cumplen con esplendidez. La rigurosidad en su cumplimiento se asegura por la apreciación racional de causa y efecto por parte del nativo combinada con cierto número de sentimientos sociales y personales tales como ambición, vanidad, orgullo, deseo de destacarse por la exhibición y también por el cariño, amistad, devoción y lealtad al allegado.

Casi es innecesario añadir que la "ley" y los "fenómenos legales" tal como los hemos descubierto, descrito y definido en una parte de Melanesia, no constituyen instituciones independientes. La ley es mejor un aspecto de su vida tribal, una fase de su estructura, que sistemas independientes, socialmente completos en sí mismos. La ley no estriba en un sistema especial de decretos que prevén y definen cualquier forma posible de su incumplimiento y que proporcionan las barreras y remedios necesarios al caso, sino que la ley es el resultado específico de la configuración de obligaciones que hacen imposible al nativo el eludir sus responsabilidades sin sufrir por ello en el futuro.

XII

SOLUCIONES LEGALES ESPECIFICAS

Las raras disputas que a veces tienen lugar toman la forma de un intercambio público de reconvenciones *(yakala)* donde las dos partes contendientes, asistidas por amigos y parientes, se encuentran, se arengan una a otra y se lanzan recriminaciones mutuas. Estos litigios permiten dar rienda suelta a los sentimientos de la gente a la par que muestran la tendencia de la opinión pública, todo lo cual ayuda a resolver las disputas. Sin embargo, a veces parece como si sólo sirviesen para endurecer más a los litigantes. En ningún caso hay un tercero que pronuncia una sentencia definida y raramente se llega a un acuerdo en aquel mismo momento. El *yakala* es por lo tanto un arreglo legal especial de pequeña importancia que en realidad no llega al corazón de la compulsión legal.

Se pueden mencionar aquí algunos otros mecanismos legales. Uno de ellos es el *kaytapaku*, la protección mágica de la propiedad por medio de maldiciones condicionales. Cuando un hombre posee cocoteros o palmeras de areca en puntos distantes donde es imposible vigilarlos, pega una hoja de palmera al tronco del árbol como indicación de que ha proferido una fórmula que automáticamente traerá desgracias, males, dolencias, etc., al ladrón. Otra institución que tiene un lado legal es la *kaytubutabu*, una especie de magia practicada sobre todos los cocoteros de una colectividad para inducir su fertilidad, por lo general cuando se aproxima una fiesta. Esta magia

lleva consigo la prohibición estricta de recoger los cocos o de participar de ellos incluso cuando son importados. Una institución similar es la *gwara* (1). Se planta un palo largo en el arrecife, que representa la prohibición de exportar ciertos objetos valiosos, intercambiados ceremonialmente en el *kula*, mientras que por el contrario, su importación es fomentada. Esto es una especie de moratoria que suspende todos los pagos sin interferencia alguna con los recibos mientras que también aspira a una acumulación de objetos valiosos ante una gran distribución ceremonial. Otro aspecto legal importante es una especie de contrato ceremonial llamado *kayasa* (2). Aquí, el jefe de una expedición, el director de una fiesta o el promotor de una aventura industrial ofrece una gran distribución ceremonial. Los que participan en ella y se benefician de su generosidad y munificencia están en la obligación de ayudar al jefe durante todo el tiempo que dura la empresa.

Todas estas instituciones, *kayasa*, *kaytapaku* y *kaytubutabu* comportan vínculos especiales. Pero ni aun éstos son exclusivamente legales. Sería un grave error tratar del tema de la ley limitándonos a una sencilla enumeración de estos pocos arreglos, cada uno de los cuales sirve un fin esepcial y cumple una función muy parcial. La incumbencia principal de la ley estriba en el mecanismo social que se encuentra en la base de todas las obligaciones verdaderas y cubre una porción muy vasta de sus costumbres, aunque ni mucho menos todas ellas, como ya sabemos.

(1) Véase la relación de esta institución en *Argonautas del Pacífico Occidental* (referencias en Indice s. v. *Gwara*). También algunas descripciones en *Melanesios* del Prof. Seligman, y en la obra del presente autor *Los Nativos de Mailu* (Trans. S. Sociedad del Sur de Australia, vol. 39) sobre la *gala o gora* entre los papue-melanesios occidentales.

(2) *Argonautas*. Véase en el Indice s. v. *Kayasa*.

XIII

CONCLUSION Y PRONOSTICO

Hasta ahora he tratado sólo de un sector de Melanesia y naturalmente las conclusiones a que he llegado tienen un alcance limitado. Estas conclusiones, sin embargo, están basadas en hechos observados por un método nuevo y han sido consideradas desde un nuevo punto de vista con objeto de estimular a otros observadores a seguir una línea de estudio similar en otras partes del mundo.

Resumamos el contraste que hay entre los puntos de vista corrientes sobre este tema y los hechos aquí presentados. En la moderna jurisprudencia antropológica se afirma universalmente que todas las costumbres son ley para el salvaje y que éste no tiene más ley que sus costumbres, al mismo tiempo que obedece automática y rígidamente todas las costumbres por pura inercia. Por lo tanto no hay ley civil ni su equivalente en las sociedades salvajes. Los únicos hechos relevantes son las ocasionales infracciones en desafío de las costumbres — los delitos —. No hay mecanismo de imposición de las reglas primitivas de conducta excepto el castigo del delito flagrante. La antropología moderna ignora y aún a veces explícitamente niega la existencia de normas sociales de ninguna clase o de motivos psicológicos que hagan obedecer al hombre primitivo y cumplir con cierta clase de costumbres por razones puramente sociales. Según Hartland y todas las demás autoridades, las sanciones religiosas, los castigos sobrenaturales, la

responsabilidad del grupo y solidaridad, el *tabú* y la magia son los principales elementos de la jurisprudencia en las sociedades salvajes.

Como he indicado más arriba, todos estos argumentos son, o bien directamente equivocados, o sólo parcialmente ciertos, o por lo menos puede decirse que colocan la realidad de la vida nativa en una falsa perspectiva. Quizás no sea ya necesario seguir arguyendo que no hay hombre, por "salvaje" y "primitivo" que sea, que actúe contra sus propios instintos, u obedezca *sin saberlo* una ley que astutamente se siente inclinado a evadir o voluntariamente a desafiar; o que actúe *espontáneamente* en forma contraria a todos sus apetitos e inclinaciones. La función fundamental de la ley es contener ciertas propensiones naturales, canalizar y dirigir los instintos humanos e imponer una conducta obligatoria no espontánea; en otras palabras, asegurar un tipo de cooperación basado en concesiones mutuas y en sacrificios orientados hacia un fin común. Una fuerza nueva, diferente de las inclinaciones innatas y espontáneas, debe estar presente para que esta tarea se lleve a cabo.

A fin de hacer que esta crítica negativa sea concluyente, hemos presentado un caso concreto para mostrar los hechos de la ley primitiva tal como son y hemos mostrado en qué consiste la naturaleza obligatoria de las reglas legales primitivas.

El melanesio de la región que aquí tratamos siente incuestionablemente el mayor respeto por las costumbres de su tribu y la tradición como tal. De este modo se puede conceder mucho a las viejas teorías del principio. Todas las reglas de su tribu, triviales o importantes, agradables o pesadas, morales o utilitarias, son consideradas por él con reverencia y sentidas como obligatorias. Pero la fuerza de la costumbre, el atractivo de la tradición por sí solos no serían bastante para contrarrestar las tentaciones del apetito, la codicia o los dictados del interés personal. La mera sanción de la tradición — el conformismo y el conservatismo del "salvaje" — opera a menudo y opera

sola en la imposición de usos y costumbres, así como en la conducta pública y privada en todos los casos donde son necesarias algunas reglas para establecer el mecanismo de la vida común y la cooperación junto con procederes ordenados — es decir, opera donde no hay necesidad de inmiscuirse con el interés personal y la inercia, o de hostigar a acciones desagradables o detener propensiones innatas.

Hay otras reglas, dictados e imperativos que requieren y tienen su tipo especial de sanciones además del mero atractivo de la tradición. Los nativos de la parte de Melanesia que hemos descrito tienen que ajustarse, por ejemplo, a un tipo de ritual religioso muy estricto, especialmente en entierros y en lutos. Hay también imperativos de conducta entre parientes. Y finalmente hay la sanción del castigo tribal ocasionado por una reacción de cólera e indignación de toda la colectividad. La vida humana, la propiedad y por último, aunque no menor en importancia, el honor personal, están salvaguardados, en una comunidad melanesia, por esta sanción del castigo tribal, lo mismo que tales instituciones como jefatura, exogamia, rango y matrimonio que desempeñan un papel principalísimo en la constitución de sus tribus.

Cada clase de reglas mencionadas se distingue de las restantes por sus sanciones y por su relación con la organización social de la tribu y de su cultura. Estas reglas no forman esa masa amorfa de usos de la tribu o "conglomerado de costumbres" del que tanto hemos oído hablar. Esta última categoría, las reglas fundamentales que salvaguardan la vida, la propiedad y la personalidad forman la clase que se puede descubrir como "ley criminal", muy a menudo exagerada por los antropólogos y falsamente asociada con el problema de "gobierno" y "autoridad central" e invariablemente arrancada de su contexto propio de otras reglas legales, ya que — y aquí llegamos al punto más importante de todos — existe una clase de reglas obligatorias que regulan la mayoría de los aspectos de la vida de la tribu y las relaciones personales entre parientes, miembros del mismo clan y de la misma tribu, que fijan

las relaciones económicas, el ejercicio del poder y de la magia, el estado legal de marido y mujer y de sus respectivas familias. Estas son las reglas de una comunidad melanesia que corresponde a nuestra ley civil.

No hay sanción religiosa a estas reglas, ni miedo, supersticioso o racional, que las haga cumplir, ni castigo tribal en caso de que alguien las infrinja, ni siquiera el estigma de la opinión pública o de la censura moral. Ahora pondremos al descubierto las fuerzas que hacen cumplir estas reglas y veremos que no son simples, pero sí claramente definibles, que no pueden describirse con una sola palabra o concepto, pero muy reales de todos modos. Las poderosas fuerzas compulsivas de la ley civil de Melanesia hay que buscarlas en la concatenación de las obligaciones, en el hecho de que están ordenadas en cadenas de servicios mutuos, un dar y tomar que se extiende sobre largos períodos de tiempo y cubre amplios aspectos de interés y actividad, añadiéndose a todo esto la forma conspicua y ceremoniosa como tienen que cumplirse la mayor parte de las obligaciones legales. Esto es efectivo porque afecta la vanidad y el amor propio, y el deseo de autoafirmación y ostentación de la gente. Resulta, pues, que el poder compulsivo de estas reglas procede de la tendencia psicológica natural por el interés personal, de la ambición y de la vanidad puestos en juego por un mecanismo social especial dentro del que se enmarcan estas acciones obligatorias.

Con una "definición mínima" de la ley, más amplia y más elástica, no hay duda alguna de que se descubrirán nuevos fenómenos legales del mismo tipo que los encontrados en el noroeste de Melanesia. No hay duda de que las costumbres no se basan sólo en una fuerza universal indiferenciada, ubicua, esta inercia mental, aunque esto existe indudablemente y añade su parte a las otras compulsiones. En todas las sociedades debe haber una clase de reglas que son demasiado prácticas para ser apoyadas por las sanciones religiosas, demasiado pesadas para ser dejadas meramente a la buena voluntad y demasiado personalmente vitales para los individuos para que

cualquier agencia abstracta pueda hacerlas cumplir. Este es el terreno de las reglas legales y me aventuro a predecir que se encontrará que la reciprocidad, la incidencia sistemática, la publicidad y la ambición serán los factores principales en la poderosa maquinaria de la ley primitiva.

Arnold Gehlen

Antropología filosófica

Del encuentro y descubrimiento del hombre por sí mismo

ediciones PAIDOS
Barcelona
Buenos Aires
México

un año entero como situaciones de aprendizaje bajo la influencia orientadora del medio circundante. La capacidad de aprendizaje del ser humano y esta influencia orientadora de su medio están, por así decirlo, incluidas en el plan de desarrollo puramente biológico, siendo el niño típica y normalmente (aunque anormalmente en comparación con el animal) sacado del cuerpo materno para ser sometido a dicha influencia. El hombre conserva por muchos años esta docilidad de sus funciones sensoriales, motoras y expresivas. Basándose en esto, el anatomista holandés Bolk ha observado que su edad adulta y, hasta cierto punto, toda su vida, está caracterizada por una notable retención de rasgos de la primera infancia, observación que he adoptado y recomendado en mi Antropología. Todos estos autores podrían concordar en las líneas generales de la concepción ya señalada, de manera que el surgimiento de una nueva orientación en la Antropología filosofía podría resultar una especie de trabajo en equipo impremeditado. En todo caso, ahora ya no parece utópica la idea de preguntarse: ¿No se podrá, basándose en las condiciones biológicas únicas, especiales y privativas del hombre, entender por qué es él un ser cultural? Estos aspectos del asunto parecen esclarecerse o ilustrarse recíprocamente. Por consiguiente, renunciando a opiniones o convicciones metafísicas que descartamos conjuntamente con el dualismo, es posible trazar una imagen del hombre. Si bien la anatomía, la psicología, la lingüística, etc., parecen ocuparse cada una de aspectos parciales de un ser complejo y muy extraordinario, hemos conseguido también cabida para algo así como una ciencia general de la cultura.

Cultura: la naturaleza transformada por la acción del hombre

Bajo este criterio la «esfera cultural» es seguramente, en una primera aproximación, el ámbito natural transformado por el hombre, el nido, por decirlo así, que el hombre se construye en el mundo. Es necesario para su vida, pues le falta la adaptación innata del animal a su medio ambiente. Por eso, la cultura de los pueblos primitivos consiste ante todo en sus armas, sus herramientas, sus chozas, sus animales domésticos, sus huertos, etc., todo lo cual es naturaleza transformada, perfeccionada, naturaleza que, reformada por una actividad inteligente, provee en todas partes los elementos, los recursos técnicos para su propia reestructuración. En el concepto de «naturaleza reformada» entran también la familia y el matrimonio, las ordenaciones sociales, que provienen de lo natural examinado a fondo y organizado como materia. Y, por último, no se exceptúa aquello que en mitología y religión sigue pareciéndole alcanzable al espíritu descifrador de enigmas del hombre.

Todas las sociedades humanas, por sencillas que sean, poseen una interpretación global del mundo y de su propio papel en ese mundo, que en última instancia se refiere a la acción. Es decir, en la medida en que el mundo *se sustrae* a la intervención del hombre, en la medida en que no ofrece asidero a su acción transformadora y creadora de utilidad, o sea, en sus estados invariables, el mundo es *interpretado* en cierto sentido, asociando a estas interpretaciones series de actos que pueden ser simbólicos. Por lo tanto, una filosofía o concepción del mundo o mitología aparece como interpretación del sentido de *los estados del mundo no susceptibles de modificación*, convirtiéndose esas interpretaciones en *motivos* para actos que serán ante todo de culto o rituales ante ese componente del mundo al cual es preciso resignarse, como ante la muerte. Este aspecto de la cultura puede relacionarse también con el hombre como ser activo, y así, de acuerdo con esta idea, podemos insertar en nuestro esquema el conjunto completo de etnología y ciencia de la cultura, prehistoria y etnografía. La precitada Antropología cultural confirma plenamente nuestra hipótesis, pues el resultado asombroso de este polifacético estudio de la cultura, practicado con tanto éxito en América, consiste en que nos da una imagen perfectamente clara de la extraordinaria plasticidad humana. Si utilizándola revisamos un par de docenas de culturas foráneas, llegaremos al sencillo enunciado: «No hay nada que no hubiera». Ése es, como quien dice, el producto abstracto de esta investigación cultural, si se realiza la conveniente extensión de la hipótesis instintiva y de la falta de fijación del hombre, como asimismo de la energía, riqueza, variedad y fantasía de su acción. Cada una de las miles de culturas primitivas representan un mundo típico, inconfundible, y no es fácil hacer afirmaciones sobre constantes innatas que excedan lo más general. La fluidez de la vida pulsional del hombre, la vivacidad de su fantasía, la variedad de circunstancias externas frente a las cuales reacciona cada vez, todo esto produce una vegetación tan exuberante que a cada paso se abren mundos nuevos. Como ejemplo accesible y sencillo, se puede indicar el libro de H. Schelsky *Sociología de la sexualidad*, en el que el autor argumenta antropológicamente, relacionando la variedad de las instituciones, la desconcertante y contradictoria abundancia de costumbres en este terreno con la falta de fijación del ser humano, con el carácter desbordante de su

vida pulsional. Parece como si uno de los poderes principales de la cultura humana consistiera en sacarles provecho, bajo el apremio de la necesidad, a las circunstancias naturales originarias encontradas. La cultura humana consiste además, esencialmente, en ordenar y *estabilizar*. Aun a riesgo de pasar por excéntrico, uno intenta encontrar algo de estabilidad y de orden en el caos al parecer siempre dispuesto en el corazón del hombre, esforzándose por salvar a través del tiempo algo de previsibilidad y continuidad. De ello se desprende ostensiblemente un segundo gran tema del concepto de cultura encuadrado también en la hipótesis de la cual partí.

El sentido de las instituciones sociales

Tocamos ahora un tema muy serio. Nietzsche habló una vez del hombre como del animal no fijado. La expresión es alarmante; significa que no existen en ese animal peculiar fijaciones concluyentes, que de por sí es un ser inestable, propenso al estado caótico, a la degeneración. Los mitos antiguos, que siempre mencionan a los dioses imponiendo al caos un orden universal, se referían a la predisposición humana a lo caótico. El estudio comparativo de la cultura y la Antropología cultural nos proporciona un resultado importantísimo: revela la inventiva y el ingenio increíbles del hombre empleados desde tiempos inmemoriales para mantener, en las condiciones más arduas e inclusive a costa de una parcialidad tremenda, instituciones y costumbres que sirvan de base para un entendimiento y como garantía de confianza mutua y de un orden no cuestionable ya. De modo que la complejidad, la parcialidad y a menudo la rareza de las instituciones humanas se pueden explicar si concebimos al hombre como el ser abandonado por los instintos. Si el hombre está abierto al mundo, si su conducta está determinada por los sucesos externos, por los nuevos datos; si el alcance instintivo es pobre e inseguro, entonces la facilidad para extraviarse pasa a constituir uno de sus rasgos principales. Se sabe que la instancia que fija al hombre directivas y puntos de estabilización es lo denominado con la palabra *moral*, cuyo designio consiste en garantizar la seguridad e inmutabilidad de las relaciones sobre una base de confianza mutua. Ya se ha mostrado que las *instituciones* de una sociedad, sus organizaciones, leyes y estilos de conducta –las formas permanentes de cooperación existentes como sistemas económico, político, social, religioso– hacen de refuerzos exteriores, de piezas de unión entre los hombres, que aseguran el lado interno de

la moral. El interior humano es un terreno demasiado escabroso para aventurarse en él. Toda la historia humana y la historia de la cultura demuestran la variabilidad de las instituciones que son apoyo y asidero externo. Pero es de suma importancia que esta variación sea paulatina. Cuando se destruyen las instituciones de un pueblo se libera toda la inseguridad elemental, la tendencia que hay en el ser humano a la degeneración y al caos. Esto lo hemos observado más de una vez, como también la analogía oculta, pero no menos inquietante, con las manifestaciones de decadencia en los pueblos primitivos, cuando los invadió la civilización europea con dinero, licores y escuelas, alterando sus normas tradicionales.

En los siglos XVIII y XIX hubo concepciones más bien idealistas del hombre, cuyo prestigio difícilmente puede desvincularse del respaldo que le ofrecía una tradición social secular, no perturbada en lo esencial, que aseguraba la buena fe de semejante optimismo. Hoy esas concepciones nos parecen ingenuas, irreales y antipolíticas.

Los confines de esta nueva Antropología se han extendido mucho, lo que en definitiva se debe a las circunstancias que anoto a continuación. Como traté de demostrarlo, se ha conseguido aunar otra vez el enfoque biológico con el de la historia de la cultura y de las ciencias culturales en general. Al desprenderse del esquema dualista y de la metafísica, la concepción del hombre como ser activo parece útil y fructífera, pues la acción es, por una parte, actividad de un organismo –de un organismo inteligente– y, por otra parte, efectúa algo en el mundo, introduce un cambio, le otorga finalidad, interviene. Así se establece la vía de enlace en que finalmente se encuentran los enfoques biológico y científico-cultural del hombre. La imagen resultante no contradice, al menos ostensiblemente, las experiencias poco inofensivas que la humanidad ha estado haciendo consigo misma siempre y especialmente en las últimas décadas. Además, la Antropología está cerca de la experiencia, no es dogmática, está abierta a los conocimientos nuevos no sólo en su aspecto empírico y filosófico ya expuesto, sino también a los de las ciencias especiales. A decir verdad, solamente en esta forma se ha podido reimportar el adelanto logrado por los americanos en las últimas décadas en materia de investigación cultural, progreso que fue esencialmente metódico y consistió en la atención imparcial a las múltiples facetas del objeto.

Estamos viviendo en una época en que el dominio de la naturaleza por el hombre plantea pocos problemas. Ha alcanzado una efectividad que ni siquiera en sus utopías pudieron imaginar cabalmente los siglos anteriores, una perfección que nos pone en el «apuro de la

riqueza». Como se señala bajo la voz «energía atómica», las dificultades que se oponen al dominio pleno de la naturaleza son ahora de orden más moral que técnico. Pero al mismo tiempo la proliferación de la especie humana, con índices de brusco crecimiento demográfico, ha llegado a constituir un problema de primera magnitud. Se puede decir que el hombre está empezando a desplazar a la naturaleza; por consiguiente, volverá los ojos hacia sí mismo, inquirirá acerca de sí mismo. He aquí otro ejemplo sorprendente de una coincidencia que siempre se puede observar: los progresos en el mundo exterior traen aparejadas tentativas por alcanzar igual nivel de dominio espiritual por reactualizar la formulación de los problemas.

2. DE LA ESENCIA DE LA EXPERIENCIA

I

Estimación diferente de la experiencia

Cuando calificamos a un individuo de pedagogo, político, soldado o marino «experimentado», estamos aplicándole la calificación máxima en esos rubros y es difícil hallarle un título más alto. Pedagogo o soldado «genial» ya sería exagerado, porque la palabra «genial» expresa una cualidad o capacidad incomparable, muy poco frecuente y casi mágica, que roza las ideas de facilidad y don natural; y porque además parece destinada al ámbito puramente espiritual, casi exclusivamente artístico. Con la palabra «experiencia», en cambio, se designa la elaboración en detalle y el dominio en todos sus aspectos de esferas vitales ricas en contenido y polifacéticas.

Éste era también el matiz de la correspondiente voz griega, *Empeiria*. Las palabras griegas *Empeiria*, *Techne* (destreza) y *Episteme* (conocimiento) indicaban práctica de muchos años, habilidad, competencia, eficiencia probada y perspicacia. Eso lo expresa también el vocablo experiencia, cuyo significado se extiende más allá de la especialización y llega hasta la experiencia en general, la experiencia de la vida. Una persona de este tipo no se ve avasallada por las múltiples exigencias que la vida nos impone normal o sorpresivamente, sino que está a la altura de ellas, y hace frente a todas las situaciones con igual decisión, inequívoca en cuanto a voluntad, y versátil en cuanto a ejecución, como procede el técnico sobresaliente en su ramo. Poseer experiencia de la vida en este sentido es infrecuente e importante, especialmente en tiempos civilizados, en que la fácil satisfacción de las necesidades esenciales hace superfluo el desarrollo de las fuerzas elementales robustecidas por los obstáculos.

mismo, escapan a la falta de interés general con que se consideran las ideas y se las olvida.

La eterna revolución contra la condición de criatura destinada a la ruda necesidad y los deberes penosos: esa eterna revolución de la cual el ser humano sale cada vez más espontáneo y temible, no habrá terminado mientras algunos grupos selectos y «minorías creadoras» no acepten el desafío poco común que hay en ese resultado lógico e imperativo, pero insensato: la tendencia a vivir bien sobre la faz de la tierra.

Desde que la civilización tomó este rumbo, el hombre está experimentando consigo mismo en un terreno donde jamás lo hiciera antes. Al tratar de sustraerse al yugo de las circunstancias, se entrega a algo que todavía no conoce bien y acerca de lo cual tiene las opiniones más frívolamente optimistas. Es decir, se entrega a sí mismo.

5. HOMBRE E INSTITUCIONES

A partir de Max Scheler, la Antropología filosófica se basó ante todo en enunciados generales acerca del hombre concebido en abstracto, pero ganó en claridad al comprobarse muchas comparaciones del ser humano con el animal. Éste sirvió, por decirlo así, de telón de fondo contra el cual se destacaba eficazmente la figura del hombre. Se estableció así con seguridad su situación única en el reino animado y, de acuerdo con una convicción más antigua y más honrosa, la diferencia específica recayó en el *espíritu*.

Pero al hacerlo la interpretación de Scheler, novedosa y dinámica en todos sus detalles, volvió a una problemática rígida y dual cuyo escaso provecho para nuevas indagaciones se había evidenciado ya desde Descartes: a un dualismo cuerpo-espíritu. Para progresar en Antropología, era preciso atacar este esquema. El arte de la investigación científica consiste a menudo en una especie de estrategia, la de prescindir deliberadamente de temas que se han revelado infecundos y que no se piensa plantear. Es preciso distribuir de nuevo las prioridades de planteamiento y dar la espalda a esas fórmulas estereotipadas. Yo he tratado de sacar este problema del espíritu de su soberanía subjetiva para incorporarlo a otro contexto donde pudiera estudiarse, por así decirlo, como predicado, esto es, en alguna forma correlativa, lo que se torna posible refiriendo los enunciados no al espíritu, sino a la conducta inteligente del hombre.

En efecto, recurriendo a describir lo estudiado objetivamente se puede demostrar que, en vista de su constitución biológica, el hombre no podría conservarse dentro de la naturaleza tal como ésta es, cruda, de primera mano; sino que debe vivir de la *transformación* –de la modificación práctica, efectiva– de cualquier realidad natural con la que se encuentre. Su actividad inteligente tiende a la modificación constructiva del mundo exterior, a causa de su insuficiencia orgánica. Así, por ejemplo, debe fabricar y elaborar él mismo las armas que le están negadas orgánicamente o, si se abre paso en regiones heladas, envolverse en la piel que a él no le crece.

Al tratar este punto, se recomienda tomar en cuenta que los pensamientos sencillos, que uno cree haber asimilado en el acto, merecen una consideración más detenida, porque la hipótesis implica varias cosás.

En primer lugar, explica por qué la especie humana puede estar en el mundo entero, mucho más propagada que cualquier especie animal. Receptivo al mundo pero pobre en instintos, desprovisto de recursos orgánicos, el hombre vive de su actividad inteligente, esto es, de la modificación de *cualesquiera* circunstancias naturales dadas, para adecuarlas a sus fines; vive, pues, hasta en el linde de estepas y desiertos que ya nada brindan a su fantasía y su habilidad. En segundo lugar, los problemas del espíritu se plantearon en forma predicativa; se habló de *conducta* inteligente, se habló del lugar. El espíritu no fue considerado sujeto respecto del cual debieran hacerse afirmaciones. En tercer lugar, por la sencillez de sus premisas esta hipótesis correspondía a los estados iniciales de la cultura humana, a los antecedentes prehistóricos. Y por último, permite ampliar los contenidos sin variar fundamentalmente el esquema. Porque a una Antropología filosófica no ha de faltarle qué decir, si se le recuerdan los últimos experimentos en el espacio exterior, la ocupación de espacios extraterrestres y las incursiones más allá del campo gravitacional: la filosofía tiene que pronunciarse acerca de ellos.

La función de descarga de las instituciones

Ya en el terreno de las investigaciones científicas de la cultura y dejando atrás las observaciones más sencillas, pronto debía presentarse una problemática conocida, a saber: la del derecho, la moral, la familia, el Estado; es decir, esa temática que Hegel tratara bajo el concepto de «espíritu objetivo». También ésta fue una formulación subjetivante, por ser el espíritu sujeto de todos los enunciados posibles; de modo que dentro de la nueva hipótesis mencionada no se podía emprender cosa alguna.

En cambio, parecía prometer concebir a los individuos abstractos de la Antropología en relaciones mutuas, dejándolos actuar unos con respecto a otros o reaccionando según las circunstancias. Entonces, en su comportamiento mutuo se decantarían y consolidarían determinadas formas o reglas, que podríamos llamar modelos conductuales rígidos. De esta manera se podrían distinguir como modelos las relaciones –jurídicas, de propiedad o de dominio–, y el tema *instituciones* reemplazaría al tema «espíritu objetivo».

Se observaron así cosas que por otro medio no hubiesen salido a luz. Se tuvo, ante todo, una visión impresionante de una de las características humanas más importantes: la reducción e inestabilización de la vida instintiva, la plasticidad y fluidez de las clases de instintos. Para establecer el nexo entre esta indefinición e imposibilidad de predicción de la conducta humana en materia de instintos y las instituciones, prefiero citar la breve fórmula de Ilse Schwidetzki en la voz «Antropología» de la enciclopedia Fischer: «En el hombre los instintos no determinan, como en el animal, distintos procesos conductuales fijos. En vez de esto, de la multiplicidad de posibles modos de conducta humanos cada cultura extrae ciertas variables y las erige en *modelos conductuales* aprobados por la sociedad y obligatorios para todos los individuos que la componen. Esos modelos conductuales civilizados, o *instituciones, liberan* al individuo de un exceso de decisiones, constituyen una guía para las innumerables impresiones y excitaciones que inundan al ser humano abierto al mundo».

Desde estos puntos de vista, las instituciones aparecen como las formas de superar tareas o circunstancias de importancia vital, así como la reproducción, la defensa o la nutrición requieren una cooperación organizada y permanente. Por el otro lado, aparecen como los poderes *estabilizadores*: son las formas que un ser inseguro e inestable por naturaleza, recargado afectivamente, encuentra para soportarse, algo en que puede confiar sea en sí mismo o en los demás. Por una parte, en estas instituciones se enfocan y persiguen en común los objetivos de la vida; por la otra, los individuos se orientan en ella hacia certezas definitivas sobre qué hacer y qué no hacer, con extraordinaria ventaja de estabilizar también su vida interior, de manera que evitan en cada oportunidad entrar en conflicto afectivo u obligarse a tomar decisiones fundamentales.

Ahora bien, una institución como la propiedad o el matrimonio le resulta al individuo un modelo suprapersonal que se le presenta y al cual se somete. En otros casos, ingresa en una institución de su oficio, en una oficina, en una fábrica, consciente de que ella ha existido por mucho tiempo y que seguirá existiendo aparte del cambio de sus integrantes. Esta temática conduce a consideraciones interesantísimas y complejas, si se quiere entender en detalle cómo se convierten las acciones humanas en algo así como autoprescripciones y luego cómo se consolidan en una reglamentación objetiva, superpuesta a ellas, que el individuo encuentra ya vigente.

Para resumir podemos decir que las formas en que los individuos conviven o colaboran, las formas en que se manifiesta la autoridad o

el contacto con lo sobrenatural, cristalizan en estructuras con peso propio –las *instituciones*– que terminan por adquirir algo así como autonomía respecto de las personas. Por regla general se puede predecir con bastante seguridad la conducta del individuo, si se conoce su ubicación dentro del conjunto social, si se sabe cuáles instituciones lo encauzan. Las exigencias de la profesión y la familia, del Estado o de las asociaciones a que se pertenece no solamente rigen nuestra conducta, afectan hasta nuestro sentido de los valores y nuestras decisiones voluntarias, los cuales prosiguen sin restricciones ni dudas –en forma espontánea, es decir, lógicamente–, sin que se nos ocurra otra posibilidad, con la fuerza persuasiva de lo natural. Referido al interior del individuo, eso representa la *bienfaisante certitude*, la bienhechora certidumbre o seguridad que constituye un alivio de importancia vital, porque esta infraestructura de hábitos internos y externos permite dedicar las energías anímicas a cosas más elevadas dejándolas disponibles para iniciativas propiamente *personales*, únicas y novedosas. En Antropología, el concepto de *personalidad* sólo puede pensarse en íntima relación con las instituciones que son las únicas que le ofrecen posibilidad de un desarrollo más refinado. Sin embargo, por personalidad no entiendo yo la soberbia exaltación de sí mismo de quienes someten en demasía a la disciplina, en realidad extraordinariamente opresora, de las grandes sociedades industriales. Quiero decir que, si bien las instituciones nos simplifican en cierto modo –acuñando y tipificando, no sólo nuestra conducta, sino también nuestro pensamiento y nuestra sensibilidad–, nos permiten reservar energías para ser una individualidad original en su medio, o sea, para actuar aportando mucho, con inventiva, con provecho. Quien quiera ser una personalidad no sólo dentro de su medio, sino en todos los medios, sólo está destinado a fracasar.

Avancemos un paso más, preguntando qué ocurre exactamente cuando se desintegran o trastornan las instituciones. Esto sucede en cada catástrofe histórica, en las revoluciones o derrumbes de estructuras estatales, de las organizaciones sociales o culturas enteras, como asimismo cuando culturas agresivas intervienen en otras más pacíficas. El efecto inmediato consiste en una *privación de la seguridad* de los individuos afectados, y ese efecto llega hasta lo más profundo: la desorientación altera los centros morales y espirituales, porque ahí está anclada la certeza de lo evidente. Así, afectado hasta los estratos esenciales, la desorientación obliga a los individuos a improvisar, a tomar decisiones sin quererlo o a lanzarse a ciegas en lo desconocido; posiblemente también a aferrarse a toda costa a algunos principios, para al menos salir a flote. Se agrega a ello la transformación afectiva de la inseguridad en angustia, en terquedad o en excitabilidad.

Todo esto recarga de control y esfuerzo resolutivo *esos* estratos del hombre donde se requiere vivir libre de conflictos, contactos facilitados por lo ya convenido, para ser capaz de encarar situaciones más valiosas. En suma, las dislocaciones en los individuos causadas por el derrumbe de sus instituciones se traducen en primitivización; su conducta recuerda los arduos esfuerzos de los sordomudos por entender. En la literatura moderna –sobre todo en la inglesa muy reciente–, hallamos por doquier una yuxtaposición de situaciones y afectos primitivos con cúmulos de reflexiones rebuscadas; también en la pintura hay esencialmente lo que Vico llamó «barbarie de la reflexión».

Desde hace muchas décadas, se ha tenido sobrada oportunidad de realizar tales experiencias en uno mismo y en experimentos masivos. Como otros de mi edad, he presenciado dos guerras, tres revoluciones y cuatro formas de Estado; y si a estas experiencias se añaden el arte y la literatura de la época, entonces ya conoce uno todas las posibilidades de deformación afectiva, desde la rigidez hasta la acomodación excesiva y el desequilibrio, desde el odio hasta el desprecio, desde la incredulidad hasta la fe ciega. También se concibe, como reacción honrosa, la necesidad e incluso la conveniencia de esa generalizada vuelta repentina a lo fácil de entender, a lo real y directamente representable, como muestra ahora la juventud.

Mas, para generalizar de nuevo, digamos que tales catástrofes sólo se han destacado a manera de cúspides en un decurso constante, ininterrumpido, que lleva dos siglos de duración y que va paralelo con el proceso mundial de industrialización, proceso que por cierto ha vuelto a definir, ha transformado o desterrado, ha destruido y desintegrado –con mayor o menor lentitud, pero igualmente a fondo– todos los moldes de vida, los ideales y criterios del mundo altamente civilizado anterior a la industrialización. Por eso, se advierte más fácilmente allí donde no se oponen al cambio resistencias masivas; es decir, en el arte y la literatura, ámbitos en los que se demolieron en pocos años normas de creación que habían regido por siglos, después de lo cual se recomenzó en todas partes, debiendo buscar cada uno su propio abecedario. El primitivismo de estas creaciones fue innegable; se continúa con él y se mantiene, así como los soldados de Castro no se cortan la barba que usaban en la selva, y todos conocemos la desorientación e indecisión del público respecto de ese arte. También en los cuadros y las obras literarias –las de Kafka, por ejemplo– se hace

evidente la pérdida del equilibrio y el tambaleo de los centros de gravedad; todo eso tiene su carácter de necesidad y su lógica intrínseca.

Así también llegó a su fin, tras 5.000 años de duración, la era de los reyes. Su sistema de instituciones, su ética, se demostraron incompatibles con las condiciones de la sociedad industrial, en la que se puede practicar cualquier ética –inclusive la del humanitarismo más elevado–, excepto la de la lucha del hombre contra el hombre y, por ende, tampoco la de la nobleza. Cada sector de la vida está variando a fondo su organización. Naturalmente, tampoco las ciencias escapan a los efectos de este prolongado trastorno. Así, por ejemplo, algo extraordinario –de lo que casi no nos dimos cuenta–, sucedió al ser desalojada la racionalidad de la filosofía, que fuera su asilo durante 2.000 años. Puede ser que hoy la racionalidad haya emigrado al proceso industrial o a algunos escritores, tal como se tiene la impresión de que el escepticismo y el estoicismo –esas grandes corrientes antiguas– están esperando en vano su Renacimiento en la filosofía. Mejor que se hubiera quedado con Gottfried Benn.

Pero con estas observaciones, necesariamente sólo indicativas, he adelantado en mi tema y por el momento voy a dejar de lado la cuestión –muy dudosa y digna de meditarse, tratada casi únicamente por los poetas anglosajones contemporáneos– de si esta desorientación sin límites no obligará algún día a callar acerca de todos los problemas más elevados, pensamiento que resuena constantemente en Samuel Beckett.

Exageración de la subjetividad

Prefiero continuar estas reflexiones preguntando qué efecto ejerce la destrucción de las instituciones –lenta, gradual o repentina, catastrófica– sobre las distintas personas que se regían por ellas; y la respuesta es indudable: el *subjetivismo*. De ningún modo quiero dar a entender con esto algo así como egoismo o egocentrismo –en el sentido corriente–, pero sí un apego tal a sí mismo que, de buenas a primeras y directamente, el individuo vive sus apropiaciones casuales, las convicciones e ideas que *él* se forma y las reacciones de *su propia* sensibilidad, como si tuviesen trascendencia más allá de su persona. Desamparado por las instituciones y devuelto a sí mismo, no puede reaccionar de otro modo que atribuyéndole validez general a lo que ha quedado de su vida interior; y esto se manifiesta hoy de un modo

natural, casi convincente en su ingenuidad. Paralelamente, corre una pretensión igualmente directa de exteriorizar lo que Benn llamó «ponderación personal». No deseo que se vea en estas consideraciones indicio alguno de ironía; sólo quisiera llamar la atención hacia este desbordamiento de la pretensión de importancia del subjetivismo, consecuencia del empobrecimiento institucional y de la confusión de normas –al igual que la indefensión y susceptibilidad de los mismos individuos–. Nunca como ahora estuvieron las personas más decididamente reducidas a las escasas reservas de sus eventuales cualidades anteriores; nunca se recurrió tanto a dichas reservas ni se estuvo, por ende, en peor situación a este respecto. Es natural que el estado de cosas aquí decrito se evidencie con mayor claridad en las esferas propiamente intelectuales, artísticas y literarias, pero creo lícito afirmarlo en general. Las susceptibilidades y los roces subjetivos se neutralizan en las instituciones que funcionan bien, porque la gente se pone de acuerdo a base de las cosas; mas si desaparecen, de ningún modo son reemplazables por la llamada discusión abierta. Resulta así la paradoja de que, mientras mayor uso hacen los individuos de la libertad fundamental de expresar sus opiniones –o sea, de exhibir sin ambages su subjetivisno–, menos contacto genuino se produce. Por el contrario, si se quiere mantener este contacto, es preciso llevar la discusión a cuestiones secundarias; la comunicación se produce cuando se elude lo esencial, y justamente por eso fracasa de nuevo: Ionesco, exagerando la situación hasta hacerla grotesca, la expresa, sin embargo, con exactitud.

Quiero ahora volver a los problemas generales y acercarme a la conclusión de mis explicaciones. La tesis que aquí he defendido –que la exaltación de la subjetividad es, como si dijéramos, el precipitado por evaporación del elemento institucional, no existiendo institucionalización de lo subjetivo–, después de exponerla en un libro en 1956, fue impugnada por Helmut Schelsky, quien sostenía que hay tanbién instituciones secundarias de esta clase –reorganizaciones, por así decirlo–, cuyo propósito sería hacer fructífera esa subjetividad inestable, versátil e inconsecuente. Ahora creo que él tiene razón y que de allí pueden derivar aclaraciones interesantes. Por ejemplo, en el arte plástico de nuestros días sin duda han desaparecido las reglas anteriores que limitaban esa rama artística. Ya no hay ideales cuya validez evidente pueda el artista dar por existente en él mismo y en el público y que el arte deba reconocer y hacer realidad; no hay ninguna sociedad dominante que los cultive como manifestación o emanación, ni reglas facultativas mantenidas por largo tiempo y enriquecidas por

muchas generaciones; ya no existe el *oficio*, ni el deseo de servir. Todo eso ha desaparecido. En su reemplazo se ha desencadenado desde hace décadas una lluvia de ocurrencias e invenciones. Toda ocurrencia es subjetiva y, por lo tanto, de un valor estético puramente casual y a menudo chocante para cualquiera que no sea su autor. Este mundo, inconsciente en alto grado, está sostenido y afirmado por una armazón de instituciones surgidas muy recientemente (50 años atrás aún no existían), algo así como una logia intercontinental establecida entre Nueva York, París y Londres, en la cual cooperan tratantes en arte, aficionados, directores de museos, coleccionistas especuladores, empresarios de exposiciones, críticos de arte, editores, etc.; es decir, un círculo excitante, donde literalmente todas las pasiones humanas encuentran oportunidad. Por consiguiente, yo diría: que ya hay una institucionalización secundaria del subjetivismo, basada, desde luego, en que la posesión de obras de arte no indica riqueza, sino que *es* riqueza. Así se explica que algunas tendencias artísticas nacidas hace décadas de la desintegración de tradiciones y de la liberación irrestricta de la subjetividad, hayan llegado a ser hoy entidades de alcance mundial con un despliegue enteramente capitalista, de una insolvencia inaudita, simplemente, porque la adquisición de originales de artistas importantes le está vedada a las personas sin muchos recursos.

Las ideas y las instituciones

Para terminar, voy a desarrollar otro argumento que deben conocer los intelectuales jóvenes, porque contradice sus concepciones. Mi tesis es que los sistemas de ideas de toda índole deben su estabilidad y su validez perdurables –inclusive su probabilidad de sobrevivir– a las instituciones en que están incorporadas. Dicho de otro modo, una conexión de pensamientos como tal, un conjunto de ideas, puede propagarse, gracias a su autoevidencia, siempre que responda a las necesidades de una época y de una cultura, pero no puede mantenerse por sus propios medios. Su idealidad tiene legitimidad en cuanto sistema jurídico y suma de normas y tradiciones jurídicas. Este conjunto legal tiene realidad como sistema estable, o sea, vigencia real, en las instancias de la vida jurídica: tribunales, autoridades administrativas, procuradurías, Facultades de Derecho, proyectos de ley parlamentarios y sección legal de las empresas industriales. Ahí reside el derecho como un conglomerado operante que se va perfeccionando y al que es posible servir sin tener que maniobrar en el movedizo terreno de lo subjetivo. Si se suprimieran dichas instituciones, seguramente subsistiría en el ser humano el sentido de justicia, pero como una entidad indigna de confianza, meramente afectiva y con pocos medios de expresarse. En materia de religión, la historia de las sectas ilustra sobradamente lo efímero de movimientos entusiastas, que existieron en su ambiente sólo por el testimonio y el poder de persecución de sus fundadores y no consiguieron constituir iglesia.

Agreguemos otro ejemplo: las numerosas ideologías socialistas que compitieron en Francia en la primera mitad del siglo XIX –la de Fourier, la de Proudhon, etc.– sólo alcanzaron la notoriedad que dispensa la literatura, a diferencia del marxismo, que de antemano apareció en Alemania como partido organizado y disciplinado.

En Alemania esto ha sido contradicho, seguramente por evidencias muy arraigadas; mas a un sociólogo no se le escapa que las ideas tienen pocas probabilidades de imponerse por sí solas. Requieren individuos que se empeñen en propagarlas, que les ayuden a abrirse paso y que a su vez coordinen entre ellos este trabajo. El mero intercambio literario entre escritor y lector sólo tiene una importancia secundaria. Hacer ver que las ideas de Rousseau o de Voltaire «se habrían divulgado en Francia» y finalmente habrían «conducido a la Revolución», es irreal, es fomentar el error. ¡Como si las fuerzas verdaderamente actuantes en la historia fuesen los escritores! Es preciso buscar siempre las asociaciones concretas que se propusieron difundir ciertas ideas, imponerlas y demostrarlas. En el caso de nuestro ejemplo, fueron los clubes repartidos por toda Francia y bien coordinados por activistas burgueses extremistas, de quienes en algunos casos –como en Dijon– sabemos hasta el nombre, oficio y modo de operar. Las ideas no solamente se comentan: *son* difundidas; sólo tienen eficacia cuando se trabaja en pro de ellas; movilizan a los individuos solamente cuando son apoyadas por otros individuos, y en este caso concreto por estos círculos detectables. No hay teoría más falsa y descarriadora que la hegeliana del autodinamismo de la idea, que sin duda favoreció considerablemente la propensión de los alemanes a vincular idealismo con irrealidad. Una filosofía empírica como la aquí expuesta –si toma la palabra experiencia en sentido exigente– llega también a conclusiones prácticas, y en último término éticas, como ésta, por ejemplo: no importa tanto discutir las ideas, como ayudarles a adquirir una legitimidad merecida y duradera.

VOLUME 87 MARCH 1974 NUMBER 5

HARVARD LAW REVIEW

THE HISTORICAL FOUNDATIONS OF MODERN CONTRACT LAW

Morton J. Horwitz *

It has long been assumed that the development of modern contract law was complete once English judges had declared late in the sixteenth century that "a promise on a promise will maintain an action upon the case." Professor Horwitz argues to the contrary that the modern will theory of contract did not appear until the late eighteenth and early nineteenth centuries, when the spread of markets forced jurists to attack equitable conceptions of exchange as inimical to emerging contract principles such as those allowing recovery of expectation damages.

MODERN contract law is fundamentally a creature of the nineteenth century. It arose in both England and America as a reaction to and criticism of the medieval tradition of substantive justice that, surprisingly, had remained a vital part of eighteenth century legal thought, especially in America. Only in the nineteenth century did judges and jurists finally reject the long-standing belief that the justification of contractual obligation is derived from the inherent justice or fairness of an exchange. In its place, they asserted for the first time that the source of the obligation of contract is the convergence of the wills of the contracting parties.

Beginning with the first English treatise on contract, Powell's *Essay Upon the Law of Contracts and Agreements* (1790), a major feature of contract writing has been its denunciation of equitable conceptions of substantive justice as undermining the "rule of law." [1] "[I]t is absolutely necessary for the advantage of the public at large," Powell wrote, "that the rights of the subject should . . . depend upon certain and fixed principles of law, and not upon rules and constructions of equity, which when applied . . . , must be arbitrary and uncertain, depending, in the extent of their application, upon the will and caprice of the

* Assistant Professor of Law, Harvard Law School. B.A., C.C.N.Y., 1959; Ph.D., Harvard, 1964; LL.B., 1967.

I wish to express my gratitude to Duncan Kennedy and Alfred S. Konefsky for many valuable suggestions.

[1] *See also* writings described in subsection II.*B.*, p. 946, *infra.*

judge." [2] The reason why equity "must be arbitrary and uncertain," Powell maintained, was that there could be no principles of substantive justice. A court of equity, for example, should not be permitted to refuse to enforce an agreement for simple "exorbitancy of price" because "it is the consent of parties alone, that fixes the just price of any thing, without reference to the nature of things themselves, or to their intrinsic value . . . . [T]herefore," he concluded, "a man is obliged in conscience to perform a contract which he has entered into, although it be a hard one . . . ." [3] The entire conceptual apparatus of modern contract doctrine — rules dealing with offer and acceptance, the evidentiary function of consideration, and especially canons of interpretation — arose to express this will theory of contract.

Powell's argument against conceptions of intrinsic value and just price reflects major changes in thought associated with the emergence of a market economy. It appears that it was only during the second half of the eighteenth century that national commodities markets began to develop in England. From that time on, "the price of grain was no longer local, but regional; this presupposed [for the first time] the almost general use of money and a wide marketability of goods." [4] In America, widespread markets in government securities arose shortly after the Revolutionary War, and an extensive internal commodities market developed around 1815.[5] The impact of these developments on both English and American contract law was profound. In a market, goods came to be thought of as fungible; the function of contracts correspondingly shifted from that of simply transferring title to a specific item to that of ensuring an expected return. Executory contracts, rare during the eighteenth century,[6] became important as instruments for "futures" agreements; formerly, the economic system had rested on immediate sale and delivery of specific property. And, most importantly, in a society in which value came to be regarded as entirely subjective and in which the only basis for assigning value was the concurrence of arbitrary individual desire, principles of substantive justice were inevitably seen as entailing an "arbitrary and uncertain" standard of value. Substantive justice, according to the earlier view, existed in order to prevent men from using the legal system in order to exploit each other. But where things have no "intrinsic value," there can be

[2] 1 J. POWELL, ESSAY UPON THE LAW OF CONTRACTS AND AGREEMENTS x (1790).

[3] 2 *id.* at 229.

[4] K. POLANYI, THE GREAT TRANSFORMATION: THE POLITICAL AND ECONOMIC ORIGINS OF OUR TIME 115 (Beacon Press ed. 1957).

[5] *See* pp. 937–41 *infra.*

[6] *See* p. 930 & note 71 *infra.*

no substantive measure of exploitation and the parties are, by definition, equal. Modern contract law was thus born staunchly proclaiming that all men are equal because all measures of inequality are illusory.

This Article will elaborate the view of the development of modern contract law outlined above. The first Section will describe the distinguishing features of the equitable conception of contract which dominated eighteenth century courts. The second Section will detail the late eighteenth and early nineteenth century disintegration of the equitable conception and the coalescence of new doctrine into the modern will theory of contract.

I. The Equitable Conception of Contract in the Eighteenth Century

The development of contract, it often has been observed, can be divided into three stages, which correspond to the history of economic and legal institutions of exchange.[7] In the first stage, all exchange is instantaneous and therefore "involves nothing corresponding to 'contract' in the Anglo-American sense of the term. Each party becomes the owner of a new thing, and his rights rest, not on a promise, but on property." [8] In a second stage, "[e]xchange first assumes a contractual aspect when it is left half-completed, so that [only] an obligation on one side remains." [9] The "third and final stage in the development occurs when the executory exchange becomes enforceable." [10] According to orthodox legal history, when English judges declare at the end of the sixteenth century that "every contract executory is an assumpsit in itself," and that "a promise against a promise will maintain an action upon the case," [11] the conception of contract as mutual promises has triumphed and, according to Plucknett, "the process is complete and the result clear . . . ." [12] "Damages were soon assessed," Ames added, "not upon the theory of reimbursement for the loss of the thing given for the promise, but upon the principle of compensation for the failure to obtain the thing promised." [13]

[7] *See, e.g.*, L. Fuller & M. Eisenberg, Basic Contract Law 121–22 (1972); F. Kessler & G. Gilmore, Contracts 27–28 (1970); T. Plucknett, A Concise History of the Common Law 643–44 (5th ed. 1956).

[8] L. Fuller & M. Eisenberg, *supra* note 7, at 121.

[9] *Id.*

[10] *Id.* at 122.

[11] T. Plucknett, *supra* note 7, at 643–44.

[12] *Id.* at 644.

[13] J. Ames, Lectures on Legal History 144–45 (1913). *See also* 3 W. Holdsworth, A History of English Law 452 (3d ed. 1923).

It is the purpose of this Section to demonstrate that, contrary to the orthodox view, the process was not complete at the end of the sixteenth century. Instead, one finds that as late as the eighteenth century contract law was still dominated by a title theory of exchange and damages were set under equitable doctrines that ultimately were to be rejected by modern contract law.

To modern eyes, the most distinctive feature of eighteenth century contract law is the subordination of contract to the law of property. In Blackstone's *Commentaries* contract appears for the first time in Book II, which is devoted entirely to the law of property. Contract is classified among such subjects as descent, purchase, and occupancy as one of the many modes of transferring title to a specific thing.[14] Contract appears for the second and last time in a chapter entitled, "Of Injuries to Personal Property." [15] In all, Blackstone's extraordinarily confused treatment of contract ideas [16] occupies only forty pages of his four volume work.

As a result of the subordination of contract to property, eighteenth century jurists endorsed a title theory of contractual exchange according to which a contract functioned to transfer title to the specific thing contracted for. Thus, Blackstone wrote that where a seller fails to deliver goods on an executory contract, "the vendee may seize the goods, or have an action against the vendor for detaining them." [17] Similarly, in the first English treatise on contract, Powell wrote of the remedy for failure to deliver stock on an executory contract as being one for specific performance.[18]

The title theory of exchange was suited to an eighteenth century society in which no extensive markets existed, and goods, therefore, were usually not thought of as being fungible. Exchange was not conceived of in terms of future monetary return, and as a result one finds that expectation damages were not recog-

[14] 2 W. Blackstone, Commentaries *440–70.

[15] 3 *id.* *154–66.

[16] *See* note 72 *infra*.

[17] 2 W. Blackstone, Commentaries *448. The title conception also appears in 1 Z. Swift, A System of the Laws of the State of Connecticut 380–81 (1796).

[18] 2 J. Powell, *supra* note 2, at 232–33. The last important appearance of the title theory in American contract law occurred in Chancellor Kent's *Commentaries*. His treatment of contracts still focused entirely upon the question of when title passes by delivery, and there was as yet no trace of a discussion of damage remedies for breach of contract. *See* 2 J. Kent, Commentaries on American Law *449–557 (1827). In a world in which markets and speculation were becoming everyday events, *see* pp. 937–41 *infra*, Kent's treatment represented the final expression of the eighteenth century view of contract as simply one mode of transferring specific property.

nized by eighteenth century courts. Only two reported eighteenth century English cases touch on the question of expectation damages for breach of contract. *Flureau v. Thornhill* [19] (1776) seems to have confronted the question of damages for the loss of a bargain. A purchaser of a lease, who sued for failure to deliver because of a defect in title, sought to recover not only his deposit, but also damages sustained as a result of the lost bargain. The report of the case does not disclose whether the plaintiff attempted to recover the increased value of the lease, or, rather, the loss he had suffered from selling stock to finance the payment. In any event, the court refused to allow him more than restitution of his payment, one judge contemptuously noting that he could not "be entitled to any damages for the fancied goodness of the bargain, which he supposes he has lost." [20]

A second English case involving the damage issue is *Dutch v. Warren* [21] (1720). The case would be irrelevant were it not for the fact that it was regularly cited by later jurists ransacking the English reports for early instances of the recognition of expectation damages.[22] The case represented a buyer's action for restitution of money paid on a stock purchase contract, the price of the stock having fallen by the time delivery was due. Although the court said the case was "well brought; not for the whole money paid, but the damages in not transferring the stock at that time," [23] the case obviously does not establish the modern rule that one may recover expectation damages in excess of the purchase price for failure to deliver stock in a rising market.

[19] 2 Black. W. 1078, 96 Eng. Rep. 635 (C.P. 1776). The decision in *Flureau v. Thornhill* may have been responsible for the widespread adoption in America during the first quarter of the nineteenth century of the rule that for breach of warranty of good title only the purchase price and not expectation damages was recoverable. *See* Horwitz, *The Transformation in the Conception of Property in American Law, 1780–1860*, 40 U. CHI. L. REV. 248, 285–88 (1973). Although the American decisions adopting this rule may reflect deeper issues concerning the relations between speculative buyers and sellers of land, they also represent the continuing influence of an eighteenth century view of contract that had not yet developed a conception of expectation damages.

[20] 2 Black, W. at 1078, 96 Eng. Rep. at 635. I have been able to find only one buyer's action for nondelivery of goods on an executory contract in the English reports of the eighteenth century, and that case did not deal with the measure of damages. In Clayton v. Andrews, 4 Burr. 2101, 98 Eng. Rep. 96 (K.B. 1767), a buyer's action for nondelivery of corn, Lord Mansfield held that the statute of frauds did not apply to executory contracts.

[21] 1 Strange 406, 93 Eng. Rep. 598 (K.B. 1760). For a more complete discussion of the case, see Moses v. Macferlan, 2 Burr. 1005, 1010–11, 97 Eng. Rep. 676, 680 (K.B. 1760) (Mansfield, J.).

[22] *See, e.g.*, Shepherd v. Johnson, 2 East. 211, 102 Eng. Rep. 350 (K.B. 1802). *See also* 1 J. POWELL, *supra* note 2, at 137–38.

[23] 1 Strange at 406, 93 Eng. Rep. at 598.

Indeed, Lord Mansfield referred to it not as establishing a rule for damages, but as illustrating the equitable nature of the action for money had and received.[24]

One of the handful of executory contracts in America before the Revolution appeared in *Boehm v. Engle* [25] (1767), in which two sellers sued a buyer who, alleging bad title, had refused to accept a deed for land. Since Pennsylvania had no equity court in which a seller could have sued for specific performance,[26] Boehm brought "a special action on the case for the consideration money" [27] or contract price, not, it should be emphasized, for the value of the lost bargain. He was thereby suing, in effect, for specific performance and not for the change in value of the land. The suit was therefore consistent with Blackstone's title theory: the contract had transferred title from seller to buyer and all that remained was an action for the price.

To appreciate the radical difference between eighteenth century and modern contract law, consider a case decided during a period in which the demise of the title theory was becoming plain. *Sands v. Taylor* [28] was an 1810 New York suit against a buyer who had received a part shipment of wheat but had refused to receive the remainder contracted for. Under the old title theory, sellers were apparently required to hold the goods until they received the contract price from the buyer. But in *Sands v. Taylor* the sellers immediately "covered" by selling the wheat in the market and thereafter suing the buyer for the difference between market and contract price. While acknowledging that there were "no adjudications in the books, which either establish or deny the rule adopted in this case," [29] the court ratified the seller's decision to "cover" and allowed him to sue for the difference. "It is a much fitter rule," it declared, "than to require . . . [the seller] to suffer the property to perish, as a condition on which his right to damages is to depend." [30] In reaching this result the court was forced to fundamentally transform the title theory. The sellers, it said, "were, by necessity . . . thus constituted trustees or agents, for the defendants . . . " [31] The trust theory was thus

[24] *See* Moses v. Macferlan, 2 Burr, 1005, 1011–12, 97 Eng. Rep. 676, 680 (K.B. 1760). *See also* Clark v. Pinney, 7 Cow. 681, 688–89 (N.Y. 1827) (*Dutch v. Warren* "most manifestly decides nothing which has a bearing upon the question of damages where the action is brought *upon the contract itself*, and not to recover back the money paid . . . .").

[25] 1 Dall. 15 (Pa. 1767).

[26] A. LAUSSAT, AN ESSAY ON EQUITY IN PENNSYLVANIA 19–27 (1826).

[27] 1 Dall. at 15.

[28] 5 Johns. 395 (N.Y. 1810).

[29] *Id.* at 406.

[30] *Id.*

[31] *Id.* at 405.

created in order to overcome a result which, though inherent in eighteenth century contract conceptions, was becoming increasingly anomalous in a nineteenth century market economy. Under an economic system in which contract was becoming regularly employed for the purpose of speculating on the price of fungible goods, the old title theory of contract, conceived of as creating a property interest in specific goods, had outlived its usefulness. As we shall see in the succeeding Section, the demise of the title theory roughly corresponded to the beginnings of organized markets and the transformation of an economic system that had used contract as simply one means of transferring specific property.

The most important aspect of the eighteenth century conception of exchange is an equitable limitation on contractual obligation. Under the modern will theory, the extent of contractual obligation depends upon the convergence of individual desires. The equitable theory, by contrast, limited and sometimes denied contractual obligation by reference to the fairness of the underlying exchange.

The most direct expression of the eighteenth century theory was the well-established doctrine that equity courts would refuse specific enforcement of any contract in which they determined that the consideration was inadequate.[32] The rule was stated by South Carolina's Chancellor Desaussure as late as 1817:[33]

> [I]t would be a great mischief to the community, and a reproach to the justice of the country, if contracts of very great inequality, obtained by fraud, or surprise, or the skillful management of intelligent men, from weakness, or inexperience, or necessity could not be examined into, and set aside.

Four years later, the Chief Justice of New York noted the still widespread opinion of American judges that equity courts would refuse to enforce a contract where the consideration was inadequate.[34]

[32] *See, e.g.*, Carberry v. Tannehill, 1 Har. & J. 224 (Md. 1801); Campbell v. Spencer, 2 Binn. 129, 133 (Pa. 1809); Clitherall v. Ogilvie, 1 Des. 250, 257 (S.C. Eq. 1792); Ward v. Webber, 1 Va. (1 Wash.) 354 (1794). On the other hand, Swift stated that "[i]nadequacy of price, abstracted from all other considerations, seems of itself to furnish no ground on which a court of equity can set aside or relieve a party to a contract." 2 Z. SWIFT, *supra* note 17, at 447–48. Swift, however, acknowledged that when inadequacy existed, together with other circumstances, a "court may conclude that the consent of the party was not free, or was conditional, thro [*sic*] mistake, fear, or misrepresentation, or under the impulse of distress, known to the other party . . . ." *Id.* at 448. In short, even according to Swift, inadequacy of consideration could lead to refusal to enforce a contract without a finding of fraud.

[33] Desaussure made this remark as an unnumbered footnote to his report of a case, Clitherall v. Ogilvie, 1 Des. 250, 259 n. (S.C. Eq. 1792).

[34] Seymour v. Delancey, 3 Cow. 445, 447 (N.Y. 1824) (Savage, C.J.).

Supervision of the fairness of contracts was not confined to courts of equity. The same function was performed at law by a substantive doctrine of consideration which allowed the jury to take into account not only whether there was consideration, but also whether it was adequate, before awarding damages. The prevailing legal theory of consideration was expressed by Chancellor Kent as late as 1822, on the very eve of the demise of the doctrine that equity would not enforce unfair bargains.[35] In contract actions at law, he wrote, where a jury determined damages for breach of contract, "relief can be afforded in damages, with a moderation agreeable to equity and good conscience, and . . . the claims and pretensions of each party can be duly attended to, and be admitted to govern the assessment." [36]

Eighteenth century American reports amply support Kent's statement. In Pennsylvania, for example, where no equity court sat,[37] eighteenth century judges instructed juries in actions on bonds that they "ought to presume every thing *to have been paid*, which . . . in equity and good conscience, *ought not to be paid*." [38] Without an equity court, Chief Justice McKean declared, courts were obliged to turn to juries for "an equitable and conscientious interpretation of the agreement of the parties." [39] As a result, Pennsylvania lawyers often argued that a plaintiff's claim on a contract "should be both legal and equitable before he can call on a jury to execute the agreement," [40] and that "[i]nadequacy of price, known to the other party, is a ground to set aside a contract."[41]

In Massachusetts, the eighteenth century rule was that a defendant in an ordinary contract case could offer evidence of inadequacy of consideration in order to reduce his damages. At three separate points in his student notes, written around 1759, John Adams indicated that "sufficient Consideration" was necessary to sustain a contract action.[42] "No Consideration, or an insufficient Consideration, a good Cause of Motion in Arrest of Judgment," Adams noted in one of these entries.[43] In *Pynchon*

[35] Kent refused to specifically enforce a contract on the grounds of the unfairness of the bargain, but he was to be overruled on appeal. Seymour v. Delancey, 3 Cow. 445 (N.Y. 1824), *rev'g* 6 Johns. Ch. 222 (N.Y. Ch. 1822).

[36] Seymour v. Delancy, 6 Johns. Ch. 222, 232 (N.Y. Ch. 1822).

[37] *See* note 26 *supra*.

[38] Holingsworth v. Ogle, 1 Dall. 257, 260 (Pa. 1788).

[39] Wharton v. Morris, 1 Dall. 125, 126 (Pa. 1785). *See also* Conrad v. Conrad, 4 Dall. 130 (Pa. 1793).

[40] Gilchreest v. Pollock, 2 Yeates 18, 19 (Pa. 1795) (argument of counsel).

[41] Armstrong v. McGhee, Addis. 261 (Pa. C.P. 1795) (argument of counsel).

[42] 1 LEGAL PAPERS OF JOHN ADAMS 9 (L. Wroth & H. Zobel eds. 1965). *See also id.* at 12, 15.

[43] *Id.* at 9. Pondering the implications of the conception of objective value

v. Brewster [44] (1766), Chief Justice Hutchinson instructed the jury in an action for a fixed price that they "might . . . if they thought it reasonable, lessen the Charges in the [plaintiff's] Account." [45] One year later Hutchinson observed that "[i]t seems hard that an Inquiring into the Consideration should be denied, and that Evidence should be refused in Diminution of Damages." [46]

Another indication of the equitable nature of damage judgments in the eighteenth century was the almost universal failure of American courts either to instruct juries in strict damage rules or else to reverse damage judgments with which they disagreed. As a result, the community's sense of fairness was often the dominant standard in contracts cases. A commentator, referring to a 1789 Connecticut commercial case, noted that "[t]he jury were the proper judges, not only of the fact but of the law that was necessarily involved in the issue . . . ." [47] Whatever they believed about the proper allocation between judge and jury on matters of law, most judges were prepared to leave the damage question to the jury. For example, in a 1786 lawsuit in which the jury's award was lower than the agreed contract price, the South

that lay at the foundation of these rules, Adams was finally undecided whether their "Inconvenience to Trade" was greater than "the Injustice" of enforcing unequal bargains.

> It is a natural, immutable Law, that the Buyer ought not to take Advantage of the sellers Necessity, to purchase at too low a Price. Suppose Money was very scarce, and a Man was under a Necessity of procuring a £ 100 within 2 Hours to satisfy an Execution, or else go to Goal. He has Quantity of Goods worth £ 500 that he would sell. He finds a Buyer who would give him £ 100 for them all, and no more. The poor Man is constrained to sell £ 500s worth for £ 100. Here the seller is wronged, tho he sell voluntarily in one sense. Yet, the Injustice, that may be done by some Mens availing them selves of their Neighbours Necessities, is not so Great as the Inconvenience to Trade would be if all Contracts were to be void which were made upon insufficient Considerations. But Q. What Damage to Trade, what Inconvenience, if all Contracts made upon insufficient Considerations were void.

1 DIARY AND AUTOBIOGRAPHY OF JOHN ADAMS 112 (L. Butterfield ed. 1961). In recognizing the inconvenience to trade, Adams had presaged later attacks on the substantive doctrine of consideration. *See* pp. 941–45 *infra*.

[44] Quincy 224 (Mass. 1766).

[45] *Id.* at 225 (emphasis deleted).

[46] Noble v. Smith, Quincy 254, 255 (Mass. 1767). The case held, by a 3–2 vote, that evidence of inadequate consideration could not be admitted in an action on a promissory note brought by the promisee against the promisor. But it is clear from the case that the court treated notes as an exception to the general rule governing contracts. Indeed, promissory notes soon became the leading example emphasized by those who wished to destroy the doctrine of consideration itself. *See* pp. 941–43 *infra*. Although Hutchinson voted to exclude evidence of inadequacy of consideration, his statement does acknowledge the general rule, which he did not contest.

[47] 1 Z. SWIFT, *supra* note 17, at 410, *referring to* Hamlin v. Fitch (Conn. Sup. Ct. Err. 1789).

Carolina Supreme Court refused to grant a new trial since "this is a case sounding in damages, and . . . the jury have thought proper to give a kind of equitable verdict between the parties . . . ." [48] Likewise, the Virginia General Court appeared to adopt the position that excessive damages were not sufficient cause for a new trial.[49] Where "positive law, and judicial precedents, [are] totally silent on the subject [of damages]," Pennsylvania's Chief Justice McKean remarked, "the principles of morality, equity, and good conscience, would furnish an adequate rule to influence and direct our judgment." [50] And it was entirely clear that it was the jury's sense of equity that would prevail. While trying a case in the United States Circuit Court, Supreme Court Justice Washington found that, by awarding a lesser judgment, the jury had ignored his instruction that the plaintiff was entitled to recover the full amount of the contract. Asked to award a new trial, Washington refused on the ground that "the question of damages . . . belonged so peculiarly to the jury, that he could not allow himself to invade their province . . . ." [51]

Further support for the existence of a substantive doctrine of consideration in the eighteenth century is found in American courts' enforcement of the rule that "a sound price warrants a sound commodity." [52] While there is no direct evidence of a substantive doctrine of consideration in eighteenth century England, several unreported trial decisions supported the "sound price" rule,[53] and as late as 1792 Blackstone's successor in the

[48] Pledger v. Wade, 1 Bay 35, 37 (S.C. 1786). *See also* Bourke v. Bulow, 1 Bay 49 (S.C. 1787).

[49] Waugh v. Bagg (Va. 1731), *reported in* 1 VIRGINIA COLONIAL DECISIONS, R77, R78 (R. Barton ed. 1909).

[50] Perit v. Wallis, 2 Dall. 252, 255 (Pa. 1796).

[51] Walker v. Smith, 4 Dall. 389, 391 (C.C.D. Pa. 1804).

[52] *See, e.g.*, Dean v. Mason, 4 Conn. 428, 434 (1822) (Chapman, J.); Baker v. Frobisher, Quincy 4 (Mass. 1762); Garretsie v. Van Ness, 1 Penning. 20, 27–29 (N.J. 1806) (Rossell, J.) (dictum); Toris v. Long, Tayl. 17 (N.C. Super. Ct. 1799); Whitefield v. McLeod, 2 Bay 380 (S.C. 1801); Mackie's Ex'r v. Davis, 2 Va. (2 Wash.) 219, 232 (1796); Waddill v. Chamberlayne (Va. 1735), *reported in* 2 VIRGINIA COLONIAL DECISIONS, *supra* note 49, at B45; 1 Z. SWIFT, *supra* note 17, at 384; *cf.* Rench v. Hile, 4 Har. & McH. 495 (Md. 1766). *See also* Z. SWIFT, DIGEST OF THE LAW OF EVIDENCE IN CIVIL AND CRIMINAL CASES AND A TREATISE ON BILLS OF EXCHANGE AND PROMISSORY NOTES 341 (1810) ("as in all other cases of the sale of personal property, our law implies a warranty"). W. Wyche's treatise on New York procedure contains an index entry, "Assumpsit for implied warranties." W. WYCHE, TREATISE ON THE PRACTICE OF THE SUPREME COURT OF JUDICATURE OF THE STATE OF NEW-YORK IN CIVIL ACTIONS 339 (1794). The text notes that the action "for deceit in selling unsound horses, or the like" was "especially of late years, usually *declared upon in assumpsit* . . . ." *Id.* at 23.

[53] *See* Parkinson v. Lee, 2 East. 314, 322, 102 Eng. Rep. 389, 392 (K.B. 1802) (Grose, J.); W. STORY, A TREATISE ON THE LAW OF CONTRACTS NOT UNDER SEAL 333 (1844); G. VERPLANCK, AN ESSAY ON THE DOCTRINE OF CONTRACTS: BEING

Vinerian Chair at Oxford, Richard Wooddeson, proclaimed the sound price doctrine to be good law.[54] Thus, one may conclude that in both England and America, when the selling price was greater than the supposed objective value of the thing bought, juries were permitted to reduce the damages in an action by the seller, and courts would enforce an implied warranty in actions by the buyer.

What we have seen of eighteenth century doctrines suggests that contract law was essentially antagonistic to the interests of commercial classes. The law did not assure a businessman the express value of his bargain, but at most its specific performance. Courts and juries did not honor business agreements on their face, but scrutinized them for the substantive equality of the exchange.

For our purposes, the most important consequence of this hostility was that contract law was insulated from the purposes of commercial transactions. Businessmen settled disputes informally among themselves when they could, referred them to a more formal process of arbitration when they could not, and relied on merchant juries to ameliorate common law rules.[55] And, finally, they endeavored to find legal forms of agreement with which to conduct business transactions free from the equalizing tendencies of courts and juries. Of these forms, the most important was the penal bond.

AN INQUIRY HOW CONTRACTS ARE AFFECTED IN LAW AND MORALS 28–29 (1825). *But see* 2 W. BLACKSTONE, COMMENTARIES *451 (warranties of good title, but not of soundness, are implied by law). An early English manuscript contract treatise had declared that an action lies on an implied warranty of merchantability, "for the party [seller] ought to make them Merchantable goods & see them well delivered without any special provision in the contract . . . ." "Of Contracts" (c. 1720) (Hargrave Ms. 265, British Museum). I am grateful to Professor John Langbein of the University of Chicago Law School for calling the manuscript to my attention. Professor Langbein believes that the eminent British lawyer, Baron Gilbert, wrote the treatise around 1720.

[54] 2 R. WOODDESON, A SYSTEMATICAL VIEW OF THE LAWS OF ENGLAND 415 (1792). On the basis of a doubtfully reported seventeenth century case, Chandelor v. Lopus, Cro. Jac. 4, 79 Eng. Rep. 3 (Ex. 1603), *noted in* 8 HARV. L. REV. 282 (1894), it was supposed by later courts that English law had never allowed an action on an implied warranty. *See, e.g.*, Seixas v. Woods, 2 Cai. R. 48 (N.Y. Sup. Ct. 1804). Let, like so many other early decisions in English legal history, the court's ruling seems to have been more the product of narrow considerations of pleading than of any direct confrontation with issues of substantive policy. *See* Hamilton, *The Ancient Maxim Caveat Emptor*, 40 YALE L.J. 1133, 1166–68 (1931); *Implied Warranty on Sale of Personal Chattels*, 12 AM. JUR. 311, 315–16 (1834). *See also* 8 W. HOLDSWORTH, *supra* note 13, at 68–70; McClain, *Implied Warranties in Sales*, 7 HARV. L. REV. 213 (1893).

[55] Various practices involving extrajudicial settlement of commercial disputes during the eighteenth century will be examined in a forthcoming book by the author.

The great advantage of the penal bond or sealed instrument was that at common law it precluded all inquiry into the adequacy of consideration for an exchange. In the medieval legal system, the use of "penal bonds with conditional defeasance," as they were called, enabled individuals to impose unlimited penalties on parties who had failed to perform agreed upon conditions.[56] The use of penal bonds declined somewhat in England during the seventeenth and eighteenth centuries as first equity, then common law courts undertook to relieve against the penal feature — the recovery of the entire sum stipulated because of even a minor breach of a specified condition.[57] Although American courts appear to have followed the English and also "chancered" these bonds,[58] virtually all large business transactions in America until the beginning of the nineteenth century took the form of two independent bonds, each of which stipulated damages for failure to perform the agreed act.[59]

Despite the practice of "chancering," the use of bonds may still have avoided an equitable inquiry into the fairness of the exchange in most cases. From the beginning of the eighteenth century English judges had begun to distinguish between penalties — which they would relieve against — and liquidated damages — which the parties were free to stipulate without the interference of courts.[60] By the time Lord Mansfield ascended to the bench, the English courts were predisposed to regard most damage provisions in bonds as liquidated and hence enforceable. "[W]here the covenant is 'to pay a particular liquidated sum,' " Mansfield declared, "a Court of Equity can not make a new covenant for a man . . . ." [61] And summing up developments during the preceding century, Lord Eldon declared in 1801 that he could "not but lament" any supposed principle that even an "enormous and excessive" damage provision, to which the parties had agreed, should be voided as a penalty.[62] "[I]t appears to me extremely difficult to apply, with propriety, the word 'excessive' to the terms in which parties choose to contract with each other. . . . It

[56] Simpson, *The Penal Bond with Conditional Defeasance*, 82 L.Q. REV. 392, 411–12 (1966).

[57] *Id.* at 415–21.

[58] *See* Wroth & Zobel, *Introduction* to 1 LEGAL PAPERS OF JOHN ADAMS, *supra* note 42, at xliii n.38.

[59] *See, e.g.*, Thompson v. Musser, 1 Dall. 458 (Pa. 1789); Cummings v. Lynn, 1 Dall. 444 (Pa. 1789); Wharton v. Morris, 1 Dall. 124 (Pa. 1785).

[60] 6 W. HOLDSWORTH, *supra* note 13, at 663 ("already equity had begun to limit . . . relief to cases in which the sum promised was clearly out of proportion to the loss incurred").

[61] Lowe v. Peers, 4 Burr. 2225, 2228, 98 Eng. Rep. 160, 162 (K.B. 1768).

[62] Astley v. Weldon, 2 B. & P. 346, 351, 126 Eng. Rep. 1318, 1321 (C.P. 1801).

has been held . . . that mere inequality is not a ground of relief . . . ." [63]

It is impossible to determine from court records whether American courts also distinguished penalties from liquidated damages. If juries were simply instructed to ignore stipulated damages in a bond and to return verdicts for actual damages, bonds could not have represented an important device for avoiding the jury's equitable inquiry into the nature of a transaction. However, it appears that even as late as the last decade of the eighteenth century the number of bonds used to effect business transactions still vastly exceeded the number of ordinary contracts containing mutual promises; this suggests that courts did not have unlimited discretion in cases involving bonds.

The late use of bonds, the absence of widespread markets, and the equitable conception of contract law conspired to retard the development of a law of executory contracts. Indeed, the primitive state of eighteenth century American contract law is underscored by the surprising fact that some American courts did not enforce executory contracts where there had been no part performance. For example, in *Muir v. Key*,[64] a Virginia case decided in 1787, a buyer of tobacco brought an action for nondelivery on a bond containing mutual promises. In the same action, the seller sued for the price. The jury returned a verdict for the buyer, which the court reversed on the ground that unless the plaintiff had paid in advance he could not sue on the contract. Thus, as late as 1787 in Virginia, there could be no buyer's action on a contract without prepayment. Nor, according to one of the judges, could the seller sue without delivery of the tobacco.[65]

[63] *Id.*

[64] St. George Tucker, Notes of Cases in the General Court, District Court & Court of Appeals in Virginia, 1786–1811, Apr. 18 & Oct. 15, 1787 (ms. in Tucker-Coleman Collection, Swen Library, College of William & Mary).

[65] Some of the language of the judges in the case may allow for other interpretations. Judge Tazewell, for example, seems to allow for enforcement of executory contracts without part performance when he states "that in an action upon mutual promise the parties may maintain reciprocal Actions . . . ." Tucker, *supra* note 64. He also may be recognizing expectation damages when he states that the jury "ought to have assessed Damages according to the differences of price, or any other *Special Damage* which plt. could have proved but here no special Damage appears: the plt. has failed in proving that he paid the whole money. The Damages therefore are excessive & a new Trial must be granted." *Id.*

At the new trial, Tucker reports, John Marshall for the defendant "submitted to the court whether the plt. must not prove paymt. on his part, in order to maintain the present Action." *Id.* The court, with Judge Tazewell dissenting, decided that he must. Thus, it is clear that enforcement of executory contracts in which there was no part performance did not yet exist in Virginia as late as 1787. Whether the requirement of payment was regarded simply as a necessary formality or whether buyers' actions were still conceived of as simply for restitution of money paid is not entirely clear.

The view that part performance was required for contractual obligation seems to have been held elsewhere in eighteenth century America as well. In his study of Massachusetts law, William Nelson states that "[a]s a general rule . . . executory contracts were not enforced . . . in pre-Revolutionary Massachusetts unless the plaintiff pleaded his own performance of his part of the bargain." [66] Thus, in his "Commonplace Book" [67] (1759) John Adams insisted that in "executory Agreements . . . the Performance of the Act is a Condition preecedent [*sic*] to the Payment." [68] For example, if two men agree on a sale of a horse, Adams wrote, "yet there is no reason that [the seller] should have an Action for the Money before the Horse is deliverd." [69] And even as late as 1795, Zephaniah Swift of Connecticut wavered between the view that performance is unnecessary for an action on a contract and the view that without either payment or delivery, "the bargain is considered of no force and does not bind either [party]." [70] It is not difficult to understand why some courts did not enforce executory contracts without part performance. The pressure to enforce such contracts would not be great in a pre-market economy where contracts for future delivery were rare,[71] and where merchants

[66] W. Nelson, The Americanization of the Common Law During the Revolutionary Era, ch. 4 at 26 (ms. of forthcoming book in author's possession).

[67] 1 LEGAL PAPERS OF JOHN ADAMS, *supra* note 42, at 4.

[68] *Id.* A modern lawyer would, of course, observe that the condition could be satisfied if the seller had tendered the horse. But in America, the first legal writer explicitly to use the concept of tender for this purpose was Daniel Chipman. *See* D. CHIPMAN, AN ESSAY ON THE LAW OF CONTRACTS FOR THE PAYMENT OF SPECIFICK ARTICLES 31–40 (1822).

[69] 1 LEGAL PAPERS OF JOHN ADAMS, *supra* note 42, at 4.

[70] 1 Z. SWIFT, *supra* note 17, at 380–81. In Gilchreest v. Pollock, 2 Yeates 18 (Pa. 1795), defendant's counsel reiterated the eighteenth century view that part performance was "a condition precedent to the payment, and the party who is to pay shall not be compelled to part with his money till the thing be performed for which he is to pay." *Id.* at 20. But by enforcing one of the early executory stock contracts the court rejected this view.

[71] Given the colonial economy, the only conceivable subject of futures contracts would have been agricultural commodities. However, "[t]he lack of a wide market for farm products was a fundamental characteristic of northern agriculture in the colonial period." P. BIDWELL & J. FALCONER, HISTORY OF AGRICULTURE IN THE NORTHERN UNITED STATES, 1620–1860, at 133 (1941). Lewis Cecil Gray states that there were "occasional instances of future-selling" in colonial Virginia. 1 L. GRAY, HISTORY OF AGRICULTURE IN THE SOUTHERN UNITED STATES TO 1860, at 426 (1941). He offers only one example, a contract entered into by George Washington with Alexandria merchants for the sale of his wheat at a uniform price over a period of seven years. And he offers no instances of futures contracts in international trade, which provided the major market for commodities during the

framed most executory transactions that did arise in terms of independent covenants through the use of bonds.

Even where executory contracts were enforced without part performance, the infrequency with which they arose slowed the development of precise legal rules for dealing with them.[72] Eighteenth century courts were regularly confronted instead with commercial cases framed in terms of penal bonds. The legal categories required to enforce independent covenants were radically different from a conception of contracts depending on mutual promises. There was no need to inquire into questions of offer and acceptance to determine whether there had been "a meeting of minds." Nor was there any reason to develop rules for regulating "order of performance" or tender where each covenant was treated as independent.[73] Finally, because of its liquidated damage provision, the bond delayed until the nineteenth century any detailed inquiry into precise rules of damages.[74]

The use of bonds seems to have substantially declined in both England and America during the early decades of the nineteenth century. If, in fact, bonds were still an important vehicle for avoiding inquiry into the fairness of an exchange during the eighteenth century, they became increasingly unnecessary as judges took control of the rules for measuring damages. Furthermore,

colonial period. Nor is any mention made of futures contracts in any other southern colonies. *See id.* at 409–33. In addition, the widespread use of bills of exchange in commercial transactions made executory contracts unnecessary. A 1791 Virginia case noted "that it was the general custom of the English merchants, who solicited tobacco consignments, to appoint agents in this country for that purpose, with power to make advances to the planters, and to draw bills [of exchange] on their principals . . . ." Hooe v. Oxley, 1 Va. (1 Wash.) 19, 23 (1791).

[72] Blackstone's confused account of assumpsit demonstrates that English lawyers had little occasion to think through the rules governing executory contracts of sale. First, he seemed to deny that executory contracts could be enforced without part performance when he wrote: "If a man agrees with another for goods at a certain price, he may not carry them away before he hath paid for them; for it is no sale without payment, unless the contrary be expressly agreed." 2 W. BLACKSTONE, COMMENTARIES *447. Order of performance and contractual obligation, it appears, are confounded. "[I]f neither the money be paid, nor the goods delivered, nor tender made, nor any subsequent agreement be entered into, it is no contract . . . ." *Id.* Here Blackstone seems to waver between part performance as a necessary requisite for contractual obligation and a conception of executory contracts made enforceable simply through tender but not delivery.

[73] The problem of order of performance, inseparably linked to the idea of executory contracts, had not been worked out until the late eighteenth century. *See* Kingston v. Preston (K.B. 1773) (Mansfield, C.J.), *summarized in* Jones v. Barkley, 2 Doug. 685, 689–92, 99 Eng. Rep. 434, 437–38 (K.B. 1781). Even after Mansfield's resolution, the problem continued to confuse American courts for another generation. *See, e.g.*, Havens v. Bush, 2 Johns. 387 (N.Y. 1807); Seers v. Fowler, 2 Johns. 272 (N.Y. 1807).

[74] *See* p. 940 & note 124 *infra*.

liquidated damage provisions were not well suited to predicting market fluctuations in an increasingly speculative economy.[75] The result was that the executory contract came gradually to supersede the bond for most nineteenth century business transactions.

Before turning to outright reversals of eighteenth century law, however, it is important to note that there was a period of uneasy compromise between the old learning and the new. The transitional nature of the late eighteenth and early nineteenth centuries is revealed most explicitly in the confused relationship between the common counts, which by the end of the eighteenth century had emancipated the law of contract from the tyranny of the older forms of action, and Blackstone's 1768 division of the field of contract law into express and implied contracts.[76]

By highlighting the express agreement, Blackstone's division was an early indication of a tendency away from an equitable and toward a will theory of contract law. It also represented an effort to create a theoretical framework as a substitute for the older forms of action. However, Blackstone himself placed the common counts in the category of implied contracts,[77] which had the significant effect of identifying them with the still dominant equitable conception of contract. Implied contracts, Blackstone wrote, "are such as reason and justice dictate, and which therefore the law presumes that every man has contracted to perform . . . ."[78] For one of the common counts — indebitatus assumpsit for money had and received — Blackstone cited Lord Mansfield's then recent path-breaking decision in *Moses v. Macferlan*,[79] in which the Chief Justice declared: "In one word, the gist of this kind of action is, that the defendant, upon the circum-

[75] *See* Graham v. Bickham, 2 Yeates 32 (Pa. 1795) (recovery allowed on a bond in excess of penalty where there had been a sharp market fluctuation).

[76] 3 W. BLACKSTONE, COMMENTARIES *154–64. Blackstone's discussion of express contracts was brief and essentially uninformative. Its most important break with the past lies in his assertion that, except for the seal, ordinary promises were "absolutely the same" as sealed instruments. 3 W. BLACKSTONE, COMMENTARIES *157. Thus, we see the beginnings of a generic conception of contracts united by common principles that transcended the particular form of action under which suits on contracts were brought. But we have yet to see any detailed elaboration of the major categories of nineteenth century contract law: offer and acceptance, consideration, and, most important, rules of contract interpretation.

[77] 3 W. BLACKSTONE, COMMENTARIES *161. He divided implied contracts into two main headings. The first group consisted of obligations imposed by courts or statutes, which arose, Blackstone thought, from an original social contract. *Id.* at *158–59. A second class, including all of the common counts, arose, he explained, "from natural reason, and the just construction of law." *Id.* at *161. In the latter class, the law assumed "that every man hath engaged to perform what his duty or justice requires." *Id.*

[78] *Id.* at *158.

[79] 2 Burr. 1005, 97 Eng. Rep. 676 (K.B. 1760).

stances of the case, is obliged by the ties of natural justice and equity to refund the money." [80]

As a result of this unrestrained identification of contract with "natural justice and equity," the triumph of the common counts threatened to reinforce the equitable conceptions which Blackstone's distinction between express and implied contracts had appeared to displace as the unifying principle of contract. This persistence of an equitable tradition in English contract law also influenced American courts. *Palfrey v. Palfrey* [81] (1772), for example, involved an action in contract by children against their mother for improper occupation of a house they had inherited on their father's death. Rejecting the defendant's argument that the proper form of action was in trespass, the Massachusetts Superior Court held that the contract action would lie. In a long and elaborate opinion, the normally form-bound and technically oriented Judge Edmund Trowbridge maintained that there was an implied contract by the defendant to pay. Judge Trowbridge noted that "it [was] necessary to know what is at this day intended by an implied contract . . . because . . . 'many of the old cases are strange & absurd, the strictness has been relaxed & is melting down in common sense of late times.'" [82] Since the plaintiffs were "clearly entitled to recover upon the merits & must in another action if not in this," the judges "ought to use [their] utmost sagacity to give them judgment . . . ." [83] Judge Trowbridge concluded, using language borrowed from Blackstone and Mansfield: [84]

> [I]t seems to be settled that implied contracts are such as reason & justice dictate; Therefore if one is under obligation from the ties of natural justice to pay another money and neglects to do it, the law gives the sufferer an action upon the case, in nature of a bill in equity to recover it; and that mere justice & equity is a sufficient foundation for this kind of equitable action.

Blackstone's interpretation of the common counts as implied contracts did not ultimately secure the dominance of the equitable conception of contract, however, because of unresolved confusions in the pleading system. It appears that the common

[80] *Id.* at 1012, 97 Eng. Rep. at 681.

[81] *Reported in* W. Cushing, Notes of Cases Decided in the Superior and Supreme Judicial Courts of Massachusetts, 1772–1789, at 1–2 & App. 1–7 (unpublished ms. in Harvard Law School Library).

[82] *Id.*, App. at 3.

[83] *Id.*, App. at 5.

[84] *Id.*, App. at 6–7. In Griffin v. Lee (Va. 1792), *reported in* Tucker, *supra* note 64, Judge Tucker protested that the common counts had been "extended far beyond the limits which appear to be reasonable" and "need[ed] no Extending."

count of indebitatus assumpsit was variously used both for suing on an express contract price and for suing on an implied contract. When it was used to sue on an express contract, another common count, quantum meruit, was employed to sue on an implied contract. "[I]n an action for work done," a mid-eighteenth century English commentator noted, "it is the best way to lay a Quantum Meruit with an Indebitatus Assumpsit. For if you fail in the proof of an express price agreed, you will recover the value." [85] As late as the turn of the century, it was also the prevailing practice in America to sue in indebitatus assumpsit for an express contract and for counts in both indebitatus and quantum meruit to be "usually joined in the declaration; so that on failure of proof of an express debt or price, the Plf. may resort ad debitum equitatis," [86] that is, to an equitable action in quantum meruit.

The transitional nature of the late eighteenth century is thus revealed in the failure of eighteenth century lawyers to perceive any latent theoretical contradictions involved in joining counts on express and implied contract.[87] Their failure to do so undoubtedly resulted from the theoretical confusions underlying the common counts themselves. Two very different conceptions of contract were submerged within actions on the common counts. One was based on an express bargain between the parties; the other derived contractual obligation from "natural justice and equity." But in the eighteenth century there was little occasion to see the two doctrinal strands as contradictory. Contract had not yet become a major subject of common law adjudication. The existence of mercantile arbitration, on one hand, and the predominance of bills of exchange, bonds, and sealed instruments in business dealings, on the other, meant that few of the legal problems that a modern lawyer would identify as contractual entered the common law courts.[88]

[85] T. Wood, An Institute of the Laws of England 555–56 (9th ed. 1763), *quoted in* C. Fifoot, History and Sources of the Common Law: Tort and Contract 363 (1949).

[86] American Precedents of Declarations 95 (B. Perham ed. 1802). In Cone v. Wetmore (Mass. 1794). (F. Dana papers, Box 16, "Court Cases A–L," Mass. Historical Society), for example, the plaintiff sued in indebitatus assumpsit for cattle sold and delivered. The Supreme Judicial Court declared that "the Deft may have every advantage of the special [express] agreement in this action which he could have had if it had been special declared on. He might show the appraised value was less than plt. demanded . . . ." *Id.* The case thus supports the proposition that a suit on the common counts could be maintained even though an express agreement existed. *Cf.* pp. 935–36 *infra.*

[87] In the nineteenth century the practice of joining counts on express and implied contracts began to be perceived as contradictory, and the rule was ultimately laid down that the existence of an express agreement precludes recovery in quantum meruit. *See* p. 952 *infra.*

[88] *See* pp. 927–31 *supra.*

In eighteenth century America, the equitable tradition in the common counts was tied not only to a general theory of natural justice but also to an economic system often based on customary prices. The striking existence of this remnant of the medieval just price theory of value can be seen in two Massachusetts colonial cases. In *Tyler v. Richards* [89] (1765), the plaintiff brought indebitatus assumpsit for boarding and schooling the defendant's son. The defendant argued that indebitatus "will not lye; they ought to have brought a *Quantum Meruit*." [90] For the plaintiffs, John Adams and Samuel Quincy argued that "[i]t ha[d] always been the Custom of this Court, to allow" the action "if the Services alledged were proved to have been done. As every Man is supposed to assume to pay the customary Price. Assumpsit is always brought for Work done by Tradesmen, and is always allowed. The Price for Boarding and Schooling is as much settled in the Country, as it is in the Town for a Yard of Cloth, or a Day's Work by a Carpenter." [91] Adams and Quincy were thus attempting to convince the court that if the value of goods or services was "settled" and bore a "customary Price," there was no difference between this action and indebitatus for a "sum certain." The defendant, however, argued that "[i]f this Proof is admitted, there will be an End of any Distinction between *Indebitatus Assumpsit* and a *Quantum meruit*." [92] The court accepted the defendant's argument and dismissed the action.

In *Pynchon v. Brewster* [93] (1766), the plaintiff brought indebitatus "upon a long Doctor's Bill for Medicines, Travel into the Country and Attendance." [94] This time, Adams, for the defendant, argued on the authority of *Tyler v. Richards* that indebitatus would not lie. The Chief Justice, however, distinguished *Tyler* on the ground that "Travel for Physicians, their Drugs and Attendance, had as fixed a Price as Goods sold by a Shopkeeper, and that it would be a great Hardship upon Physicians to oblige them to lay a *Quantum Meruit*." [95]

What emerges from these cases is that in America suits in indebitatus were sometimes based on a system of fixed and customary prices. Though the *Richards* court denied the analogy between the price of schooling and the "settled" price for a yard of cloth, it never challenged Adams' premise that the prices of most goods and services were conceived of as "settled." Similarly,

[89] Quincy 195 (Mass. 1765).
[90] *Id.*
[91] *Id.* at 195–96.
[92] *Id.* at 196.
[93] Quincy 224 (Mass. 1766).
[94] *Id.*
[95] *Id.*

while acknowledging the "uncertain" price of schooling, Chief Justice Hutchinson had no doubt that the price of a doctor's medicine and services "had as fixed a Price as Goods sold by a Shopkeeper." [96]

Of course, there could not have been a customary rate for every exchange that might be entered into and sued upon; the jury's power to set a reasonable price in quantum meruit was necessary to fill in the gaps. Indeed, it appears that the jury had discretion to mitigate or enlarge the damages even in indebitatus actions.[97] But the concept of customary prices formed the necessary foundation for a legal system which awarded contract damages according to measures of fairness independent of the terms agreed to by the contracting parties. By the end of the eighteenth century, however, the development of extensive markets undermined this system of customary prices and radically transformed the role of contract in an increasingly commercial society.

II. The Rise of a Market Economy and the Development of the Will Theory of Contract

A. Early Attacks on Eighteenth Century Contract Doctrine

For a variety of reasons, it is appropriate to correlate the emergence of the modern law of contract with the first recognition of expectation damages. Executory sales contracts assume a central place in the economic system only when they begin to be used as instruments for "futures" agreements; to accommodate the market function of such agreements the law must grant the contracting parties their expected return. Thus, the recognition of expectation damages marks the rise of the executory contract as an important part of English and American law. Furthermore, the moment at which courts focus on expectation damages, rather

[96] In *Pynchon*, Hutchinson also remarked that it was not the practice in England to allow an indebitatus for a customary price. Quincy at 224.

[97] *See* Pynchon v. Brewster, Quincy 224, 225 (Mass. 1766) (Hutchinson, C.J.); 1 Legal Papers of John Adams, *supra* note 42, at 16. This concession to jury discretion may not, however, mean that courts had eroded every practical difference between quantum meruit and indebitatus assumpsit. It was one thing to acknowledge a complete jury power to set "reasonable" prices in quantum meruit; it was another to place a special burden on the jury to modify a fixed price that the court had established as the standard measuring rod for actions in indebitatus assumpsit. In any case, all of this home-grown lawmaking was swept aside in Glover v. LeTestue, Quincy 225 n.1 (Mass. 1770), where the Massachusetts court, after hearing extensive citations of English authority, held that only quantum meruit and not indebitatus assumpsit would lie for "Visits, Bleeding [or] Medicines" by a doctor. *Id.* at 226. *Cf.* note 96 *supra*.

than restitution or specific performance to give a remedy for nondelivery, is precisely the time at which contract law begins to separate itself from property. It is at this point that contract begins to be understood not as transferring the title of particular property, but as creating an expected return. Contract then becomes an instrument for protecting against changes in supply and price in a market economy.

The first recognition of expectation damages appeared after 1790 in both England and America in cases involving speculation in stock. Jurists initially attempted to encompass these cases within traditional legal categories. Thus, Lord Mansfield in 1770 referred to a speculative interest in stock as "a new species of property, arisen within the compass of a few years." [98] In 1789 the Connecticut Supreme Court of Errors held that recovery of expectation damages on a contract of stock speculation would be usurious.[99] And as late as 1790, John Powell concluded that specific performance, and not an action for damages, was the proper remedy for failure to deliver stock on a rising market.[100]

These efforts to encompass contracts of stock speculation within the old title theory were soon to be abandoned, however. Between 1799 and 1810 a number of English cases applied the rule of expectation damages for failure to deliver stock on a rising market.[101] In America the transformation occurred a decade earlier, in response to an active "futures" market for speculation in state securities which rapidly developed after the Revolutionary War in anticipation of the assumption of state debts by the new national government. The earliest cases allowing expectation damages on contracts of stock speculation appeared in South Carolina, Virginia, and Pennsylvania.

In South Carolina, three cases between 1790 and 1794 established the rule of expectation damages in stock cases. The first case, *Davis v. Richardson* [102] (1790), involved a "short sale" of South Carolina indents, or government stock. The defendant had borrowed the stock, promising its return with interest at a future

[98] Nightingal v. Devisme, 5 Burr. 2589, 2592, 98 Eng. Rep. 361, 363 (K.B. 1770).

[99] Fitch v. Hamlin (Conn. Sup. Ct. Err. 1789), *reported in* 1 Z. SWIFT, *supra* note 17, at 410–12.

[100] 2 J. POWELL, *supra* note 2, at 232–33.

[101] The leading case is Shepherd v. Johnson, 2 East. 211, 102 Eng. Rep. 349 (K.B. 1802). *See also* M'Arthur v. Seaforth, 2 Taunt. 257, 127 Eng. Rep. 1076 (C.P. 1810); Payne v. Burke (C.P. 1799), *discussed at* 2 East. 212 n.(a), 102 Eng. Rep. 350 n.(a). While these cases deal explicitly with the question of whether damages should be measured as of the promised date of delivery or as of the date of trial, they are nevertheless also the first cases that recognize any measure of expectation damages.

[102] 1 Bay 105 (S.C. 1790).

time. "[I]n consequence of the prospect of the adoption of the funding system by Congress," the value of the stock increased and the defendant could only "cover" at a substantially higher price.[103] The South Carolina Supreme Court made no effort to conceal the significance of the damage question before it. "[I]t is of extensive importance to the community, that the principle should *now be settled and ascertained with precision*," [104] the court declared. "A great number of contracts in every part of the state, depend upon the determination of this question: and it is fortunate, that so respectable a jury are convened for the purpose of fixing a standard for future decisions." [105] And with the aid of advice from a "respectable" merchant jury, the court announced its holding: "Whenever a contract is entered into for the delivery of a specific article, the value of that article, at the time fixed for delivery, is the sum a plaintiff ought to recover." [106]

It is entirely possible, of course, that the defendant in *Davis v. Richardson* was not the stock speculator that I have supposed him to be. In specie-scarce postrevolutionary South Carolina, where bonds and securities were regularly used for money, he may simply have been treating the indents as currency. As a result, he may have been one of the earliest casualties of the almost instant creation of a speculative market for state securities after the establishment of the national government. Prevailing economic and legal conceptions about the true nature of stock transactions were in a state of flux. Twice in the next four years, lawsuits [107] involving expectation damages on stock were carried to the Supreme Court of South Carolina in an attempt to reverse the ruling in *Davis v. Richardson*. The major argument put forth by Charles Pinckney, the leader of the South Carolina bar, was that the allowance of expectation damages was nothing more than the allowance of usury.[108] In *Atkinson v. Scott* [109] (1793), where the disputed securities had appreciated by 850% in one year, the Supreme Court admitted that such contracts "must strike every mind at the first blush" as "evidently usurious." [110] If, Pinckney argued, South Carolina stock was to be treated as money, the borrower could only be expected to pay the value at the time of the contract plus interest. But in a world in which a "respect-

[103] *Id.*

[104] *Id.* at 106.

[105] *Id.*

[106] *Id.*

[107] Wiggs v. Garden, 1 Bay 357 (S.C. 1794); Atkinson v. Scott, 1 Bay 307 (S.C. 1793).

[108] Atkinson v. Scott, 1 Bay 307 (S.C. 1793) (argument of counsel).

[109] 1 Bay 307 (S.C. 1793).

[110] *Id.* at 309. *Cf.* p. 937 & note 99 *supra*.

able" jury of merchants had recognized that stocks were traded on speculation, it made no sense for courts to deny the speculative purpose of the transaction. The result was that Pinckney's argument was rejected, and by 1794, the South Carolina legal system applied the rule of expectation damages to what appear to be the first organized markets that had developed in that state.[111]

In Virginia, the transformation of legal conceptions took an identical path. In *Groves v. Graves* [112] (1790), the rule of expectation damages arose in connection with a buyer's action for securities. After a jury had awarded the plaintiff expectation damages, however, Chancellor Wythe, still reflecting eighteenth century moral and legal conceptions, enjoined the enforcement of the judgment on the grounds that the transaction "appeared to have been designed to secure unconscionable profit . . . and to have been obtained from one whom he had cause to believe at that time to be needy . . . ." [113] He allowed damages only to the extent of the original value plus interest.[114] But the Virginia Court of Appeals reversed his decree, holding that "the contract was neither usurious, or so unconscionable as to be set aside . . . ." [115] And, in marked contrast to the earlier practice of not reviewing jury damage awards,[116] the court held that the jury erred in measuring damages as of the time of trial and not as of the time of delivery.[117] The case thus suggests that judicial supervision of juries' damage awards may have arisen simultaneously with the recognition of expectation damages.

The first published opinion in Pennsylvania allowing expectation damages for failure to deliver stock certificates on a rising market was decided in 1791.[118] The rule was elaborated in a 1795 case, *Gilchreest v. Pollock*,[119] where a seller of stock sued the buyer's surety for failure to accept the transfer of United States securities that had fallen in price after the contract was made.

[111] *See* cases cited note 107 *supra*.

[112] 1 Va. (1 Wash.) 1 (1790).

[113] *Id.* at 3 (recitation of chancellor's opinion).

[114] *Id.*

[115] *Id.*

[116] *See* pp. 925–26 *supra*.

[117] 1 Va. (1 Wash.) at 4. This issue remained unsettled ten years later. In Kirtley v. Banks (Va. 1800), *reported in* Tucker, *supra* note 64 (Dec. 9, 1800), a suit for failure to deliver securities, the court instructed the jury that it "may take the price at either period, but not any higher price at any intermediate period." *Id.* The jury selected the time of delivery as its standard.

[118] Marshall v. Campbell, 1 Yeates 36 (Pa. 1791).

[119] 2 Yeates 18 (Pa. 1795). Two other cases also granted expectation damages to enforce contracts for the sale of United States securities. *See* Livingston v. Swanwick, 2 Dall. 300 (C.C.D. Pa. 1793); Graham v. Bickham, 4 Dall. 149 (Pa. 1796).

While the merchant jury in South Carolina had had no difficulty in reaching their result, the Pennsylvania court felt compelled to charge its lay jurors that "[t]he sale of stock is neither unlawful nor immoral. It is confessed, that an inordinate spirit of speculation approaches to gaming and tends to corrupt the morals of the people. When the public mind is thus affected, it becomes the legislature to interpose." [120]

The early Pennsylvania case is somewhat anomalous in that it rested on an unpublished opinion, rendered in 1786, which recognized a market price for wheat and announced that "[t]he rule or measure of damages in such cases is to give the difference between the price contracted for and the price at the time of delivery." [121] With this one exception, however, the evidence appears to indicate that the rule for expectation damages first arose in connection with stock speculation both in England and in America.[122] In England the principle of expectation damages was not generalized in cases dealing with sales of commodities until 1825,[123] and Chitty's treatise on contracts, published in 1826, is the first to announce a general rule of expectation damages for failure to deliver goods.[124]

[120] 2 Yeates at 21.

[121] Lewis v. Carradan (Pa. 1786), *cited in* 1 Yeates at 37.

[122] The first reported case in Massachusetts involving the measure of damages for nondelivery is also a securities case. Gray v. Portland Bank, 3 Mass. 364, 382, 390–91 (1807).

[123] Greening v. Wilkinson, 1 Car. & P. 625, 171 Eng. Rep. 1344 (K.B. 1825); Gainsford v. Carroll, 2 B. & C. 624, 107 Eng. Rep. 516 (K.B. 1824); Leigh v. Paterson, 8 Taunt. 540, 129 Eng. Rep. 493 (C.P. 1818).

[124] J. CHITTY, A PRACTICAL TREATISE ON THE LAW OF CONTRACTS, NOT UNDER SEAL 132 (1826). Powell's *Essay Upon the Law of Contracts* (1790) does not appear to deal with sales. His only recognition of the effect of changes in the market on contracts of sale is his statement that if, after a contract for delivery of corn, the price *falls* to 5 pounds, the buyer "will be entitled either to . . . [the] corn, or five pounds." 1 J. POWELL, *supra* note 2, at 409. He also states the rule that "if one of the parties fail in his part of the agreement, he shall pay the other party such damages as he has sustained by such neglect or refusal." *Id.* at 137. Powell cited the famous case of Dutch v. Warren, 1 Strange 406, 93 Eng. Rep. 598 (K.B. 1720), which, as we have seen, was simply an action for restitution. *See* pp. 921–22 *supra*.

In Samuel Comyn's *Treatise on Contracts*, an entire chapter is devoted to contracts for the sale of goods. While Comyn does recognize executory contracts, most of the discussion is devoted either to formation of binding contracts or to sellers' remedies for breach. In his very brief reference to buyers' actions for nondelivery, Comyn concluded only that if the buyer tenders payment, he "may take and recover the things." 2 S. COMYN, TREATISE ON CONTRACTS 212 (1807). For this conclusion he cites only an obscure early seventeenth century treatise. Indeed, this discussion is more in line with Blackstone's title theory analysis of contract as one mode of transfer of property than with a nineteenth century market approach.

Finally, with Joseph Chitty's *Treatise on Contracts*, the rule of expectation

In America the application of expectation damages to commodities contracts correlates with the development of extensive internal commodities markets around 1815. The leading case is *Shepherd v. Hampton*[125] (1818), in which the Supreme Court held that the measure of damages for failure to deliver cotton was the difference between the contract price and the market price at the time of delivery. Within the next decade a number of courts worked out the problems of computing expectation damages for commodities contracts,[126] one of them noting that "[m]ost of the [prior] cases in which this principle has been adopted, have grown out of contracts for the delivery and replacing of stock . . . ." [127]

The absorption of commodities transactions into contract law is a major step in the development of a modern law of contracts. As a result of the growth of extensive markets, "futures" contracts became a normal device either to insure against fluctuations in supply and price or simply to speculate. And as a consequence, judges and jurists began to reject eighteenth century legal rules which reflected an underlying conception of contract as fair exchange.

It has already been noted that in the eighteenth century, commercial classes endeavored to cast their transactions in legal forms which avoided the equalizing tendencies of early contract doctrine. Not surprisingly, the first direct assault upon the equitable conception of contract appeared in adjudications involving one of these forms, the negotiable instrument.

During the second half of the eighteenth century, a movement developed to eliminate the substantive significance of the doctrine of consideration in cases involving negotiable instruments. In

damages is announced: "In an action of assumpsit, for not delivering goods upon a given day, the measure of damages is the difference between the contract price, and that which goods of a similar quality and description, bore on or about the day, when the goods ought to have been delivered." J. CHITTY, *supra*, at 131–32. Interestingly, he cites only two cases decided in the previous five years.

The chapters on damages in the treatises of Powell, Comyn, and Chitty do not mention the problem of expectation damages. Rather, they address themselves exclusively to the problem of how to distinguish penal clauses from clauses providing for liquidated damages. This emphasis reveals the extent to which commercial transactions were still far more dependent on the use of bonds than on contracts. *See* pp. 927–29 *supra*.

[125] 16 U.S. (3 Wheat.) 200 (1818). McAllister v. Douglass & Mandeville, 15 F. Cas. 1203 (No. 8657) (C.C.D.D.C. 1805), *aff'd*, 7 U.S. (3 Cranch) 298 (1806), superficially resembles *Shepherd*, but there was no agreed upon contract price.

[126] *See, e.g.*, West v. Wentworth, 3 Cow. 82 (N.Y. Sup. Ct. 1824) (salt); Merryman v. Criddle, 18 Va. (4 Muni.) 542 (1815) (corn).

[127] Clark v. Pinney, 7 Cow. 681, 687 (N.Y. Sup. Ct. 1827). One earlier case involving a commodity was Sands v. Taylor, 5 Johns. 395 (N.Y. Sup. Ct. 1810), *discussed* pp. 922–23 *supra*.

1767, the Massachusetts Superior Court held by a 3–2 vote that even in an action between the original parties to a promissory note, the promisor could not offer evidence of inadequate consideration in mitigation of damages.[128] "People," Chief Justice Hutchinson declared, "think themselves quite safe in taking a Note for the Sum due, and reasonably suppose all Necessity of keeping the Evidence of the Consideration at an End; it would be big with Mischief to oblige People to stand always prepared to contest Evidence that might be offered to the Sufficiency of the Consideration. This would be doubly strong in Favour of an Indorsee." [129]

It was one thing to argue that in order to make notes negotiable a subsequent indorsee would be allowed to recover on a note regardless of the consideration between the original parties. This argument, of course, itself entailed a sacrifice of judicial control over bargains that commercial convenience was beginning to demand. It was, however, quite a different matter to exclude evidence of consideration between the original parties to the note, as the Massachusetts court decided. With this decision, it became possible for merchants to exclude the question of the equality of a bargain by transacting their business through promissory notes.

The Massachusetts decision was handed down two years after Lord Mansfield's dramatic but unsuccessful attempt to destroy the doctrine of consideration in the case of *Pillans v. Van Mierop*,[130] a case between merchants involving a promise to accept a bill of exchange. "I take it," Mansfield declared in dictum, "that the ancient notion about the want of consideration was for the sake of evidence only; for when it is reduced into writing, as in covenants, specialties, bonds, etc., there was no objection to the want of consideration." [131] While it is impossible to know from this pronouncement whether Mansfield's *ratio decidendi* was that consideration was unnecessary for all written instruments or merely for those between merchants, two conclusions are clear. First, by explaining the requirement of consideration exclusively in terms of its evidentiary value in proving the existence of a contract, Mansfield had cut the heart out of the traditional equalizing function of consideration. Second, whether or not upon reflection Mansfield would have extended these views to cover all written

[128] Noble v. Smith, Quincy 254 (Mass. 1767).

[129] *Id.* at 255.

[130] 3 Burr. 1663, 97 Eng. Rep. 1035 (K.B. 1765). There is no citation of this case in *Noble v. Smith*. The third volume of Burrow's reports was first published in 1771, four years after *Noble v. Smith* was decided.

[131] *Id.* at 1669, 97 Eng. Rep. at 1038.

instruments — where the writing was itself sufficient evidence of a contract — he at least meant to apply the rule to negotiable instruments. Indeed, as Mansfield's decision was being announced, the second volume of Blackstone's *Commentaries* was at the press, also propounding the rule that evidence of lack of consideration would not be admitted in an action on a negotiable instrument.[132] For thirteen years, English law stood thus on the verge of rejecting the ancient requirement of consideration. But in *Rann v. Hughes* [133] (1778), the House of Lords reaffirmed the requirement of consideration for written instruments.

The views of Mansfield and Blackstone were to have a greater effect than the decision by the House of Lords, however. The report of that decision was unpublished until 1800, and was unknown by American judges before the early years of the nineteenth century.[134] Thus, even after Mansfield's opinion was overruled we find Zephaniah Swift, the first American treatise writer, stating that the principle that had emerged from negotiable instruments law — he cited Blackstone — "clearly destroys all distinction between sealed and unsealed contracts." [135] The result, he concluded, was that a written contract "precludes an enquiry into the consideration." [136] A more important factor than the accident of reporting, however, was the congeniality of Mansfield's and Blackstone's views to American judges, whose own opinions were gradually inclining towards a conception of contract as a sacred bargain between private parties.

The most persistent American advocate of the Mansfield position was the able judge of the New York Supreme Court, Brockholst Livingston, whose commercial law practice before he ascended to the bench was probably second only to that of Alexander Hamilton. In 1804, Livingston reiterated the position that as between even the original parties to a negotiable instrument, the failure of consideration could not be shown. "It is not necessary, as in other simple contracts, to state a consideration in the declaration; the instrument itself imports one, and in this respect

[132] 2 W. BLACKSTONE, COMMENTARIES *446.

[133] 7 T.R. 350 n.1, 101 Eng. Rep. 1014 n.1 (1778).

[134] In the typically chaotic fashion of law reporting of the time, the decision was casually included as a footnote to the report of another case, Mitchinson v. Hewson, 7 T.R. 350, 101 Eng. Rep. 1014 (1797). The earliest recognition of the House of Lords decision in America that I am aware of is St. George Tucker's citation in his 1803 edition of Blackstone. 3 BLACKSTONE'S COMMENTARIES *446 n.1 (St. G. Tucker ed. 1803). In 1804, William Cranch acknowledged that he just learned of the decision as he was about to publish his elaborate essay on negotiable instruments. 5 U.S. (1 Cranch) 445 n.1.

[135] 1 Z. SWIFT, *supra* note 17, at 373.

[136] *Id.* *See also* Z. SWIFT, DIGEST, *supra* note 52, at 339.

partakes of the quality of a speciality [sealed instrument]." [137] Livingston extended the argument to cover simple contracts in a case decided one year later. In *Lansing v. McKillip*,[138] he dissented from the court's opinion requiring that consideration be proved by the plaintiff before he could recover on a contract. At first, he urged only that the traditional burden of proof be altered so that a defendant who wished to negate a contract be required to show lack of consideration.[139] In the process, however, he was moved to attack the very requirement of consideration itself. Ridiculing a rule of consideration that "does not demand an absolute equivalent, but is satisfied, in many cases, with the most trifling ground that can be imagined," he urged the court to "be content in point of evidence, with a declaration . . . that he has received a *valuable one*, without indulging the useless curiosity of prying further into the transaction." [140] Livingston was fully aware that his opinion directly attacked the traditional equalizing function of consideration. "Why," he asked, is a court "so very careful of a defendant's rights as not to suppose him capable of judging for himself, what was an adequate value for his promise? Would it not be more just, and better promote the ends of justice, that one, who had signed an instrument of this kind, should, without further proof, be compelled to perform it, unless he could impeach the validity on other grounds?" [141]

Like Mansfield's earlier effort, this attack on consideration initially failed, but in its most important respect it ultimately succeeded. It was part of a movement, which had begun in England during Mansfield's tenure and continued throughout the nineteenth century, toward overthrowing the traditional role of courts in regulating the equity of agreements. The underlying logic of the attack on a substantive doctrine of consideration came to fruition in America with the great New York case of *Seymour v. Delancy* [142] (1824), in which a sharply divided High Court of Errors reversed a decision of Chancellor Kent, who had refused to specifically enforce a land contract on the ground of gross inadequacy of consideration between the parties. "Every member of this Court," the majority opinion noted, "must be well aware how much property is held by contract; that purchases are constantly made upon speculation; that the value of real estate

[137] Livingston v. Hastie, 2 Cai. R. 246, 247 (N.Y. Sup. Ct. 1804).

[138] 3 Cai. R. 286 (N.Y. Sup. Ct. 1805).

[139] *Id.* at 289–91 (Livingston, J., dissenting). This position was also adopted by another judge, William Cranch. *See* 1 Cranch 445.

[140] 3 Cai. R. at 290.

[141] *Id.*

[142] 3 Cow. 445 (N.Y. 1824), *rev'g* 6 Johns. Ch. 222 (N.Y. Ch. 1822).

is fluctuating . . . ." [143] The result was that there "exists an honest difference of opinion in regard to any bargain, as to its being a beneficial one, or not." [144] The court held that only where the inadequacy of price was itself evidence of fraud would it interfere with the execution of private contracts.[145]

The nineteenth century departure from the equitable conception of contract is particularly obvious in the rapid adoption of the doctrine of caveat emptor. It has already been noted that, despite the supposed ancient lineage of caveat emptor, eighteenth century English and American courts embraced the doctrine that "a sound price warrants a sound commodity." [146] It was only after Lord Mansfield declared in 1778, in one of those casual asides that seem to have been so influential in forging the history of the common law, that the only basis for an action for breach of warranty was an express contract,[147] that the foundation was laid for reconsidering whether an action for breach of an implied warranty would lie. In 1802 the English courts finally considered the policies behind such an action, deciding that no suit on an implied warranty would be allowed.[148] Two years later, in the leading American case of *Seixas v. Woods*,[149] the New York Supreme Court, relying on a doubtfully reported seventeenth century English case,[150] also held that there could be no recovery against a merchant who could not be proved knowingly to have sold defective goods. Other American jurisdictions quickly fell into line.[151]

While the rule of caveat emptor established in *Seixas v. Woods* seems to be the result of one of those frequent accidents of historical misunderstanding, this is hardly sufficient to account for the widespread acceptance of the doctrine of caveat emptor elsewhere

[143] *Id.* at 533.

[144] *Id.*

[145] *Id.* The year before *Seymour v. Delancey* was decided, Nathan Dane had already anticipated its main thrust. *See* p. 950 *infra*.

[146] *See* pp. 926–27 *supra*.

[147] Stuart v. Wilkins, 1 Doug. 18, 20, 99 Eng. Rep. 15, 16 (K.B. 1778). Though Mansfield was laying the foundation for the subsequent rejection of the sound price doctrine, his purpose was not clearly understood. Nathan Dane, for one, misread the case as upholding the doctrine and, therefore, attempted in 1823 to show that it was "contrary to most of the settled cases in the books . . . ." 2 N. DANE, A GENERAL ABRIDGEMENT AND DIGEST OF AMERICAN LAW 542 (1823).

[148] Parkinson v. Lee, 2 East. 314, 102 Eng. Rep. 389 (K.B. 1802).

[149] 2 Cai. R. 48 (N.Y. Sup. Ct. 1804).

[150] The case relied upon was Chandelor v. Lopus, Cro. Jac. 4, 79 Eng. Rep. 3 (Ex. 1603); *see* note 54 *supra*.

[151] *See, e.g.*, The Monte Allegre, 22 U.S. (9 Wheat.) 616 (1824); Dean v. Mason, 4 Conn. 428 (1822); Bradford v. Manly, 13 Mass. 139 (1816); Curcier v. Pennock, 14 S. & R. 51 (Pa. 1826); Wilson v. Shackleford, 25 Va. (4 Rand.) 5 (1826).

in America. Nor are the demands of a market economy a sufficient cause. Although the sound price doctrine was attacked on the ground that there "is no standard to determine whether the vendee has paid a *sound* price," [152] the most consistent legal theorist of the market economy, Gulian Verplanck, devoted his impressive analytical talents to an elaborate critique of the doctrine of caveat emptor.[153] The sudden and complete substitution of caveat emptor in place of the sound price doctrine must therefore be understood as a dramatic overthrow of an important element of the eighteenth century's equitable conception of contract.[154]

B. The Synthesis of the Will Theory of Contract

The development of extensive markets at the turn of the century contributed to a substantial erosion of belief in theories of objective value and just price. Markets for future delivery of goods

[152] Dean v. Mason, 4 Conn. 428, 434–35 (1822) (Chapman, J.).

[153] *See* p. 948 *infra*. I have not meant to assert that caveat emptor is more conducive to a market economy than the contrary doctrine of caveat venditor, though this might be independently demonstrated. Rather, I have argued that the importance of caveat emptor lies in its overthrow of both the sound price doctrine and the latter's underlying conception of objective value.

[154] We can best see the nature of the attack on the "sound price" doctrine in South Carolina, the only state in which it persisted well into the nineteenth century. Urging reversal of the sound price doctrine and adoption in its place of a rule of caveat emptor, the Attorney General of South Carolina argued in 1802 that "[s]uch a doctrine . . . if once admitted in the formation of contracts, would leave no room for the exercise of judgment or discretion, but would destroy all free agency; every transaction between man and man must be weighed in the balance like the precious metals, and if found wanting in . . . adequacy, must be made good to the uttermost farthing . . . ." Whitefield v. McLeod, 2 Bay 380, 382 (S.C. 1802) (argument of counsel). If a court should refuse to enforce a contract made by a man who has had "an equal knowledge of all the circumstances" as well as "an opportunity of informing himself, and the means of procuring information . . . ," he maintained, "good faith and mutual confidence would be at an end. . . . To suffer such a man to get rid of such a contract, under all these circumstances," he concluded, "would establish a principle which would undermine and blow up every contract . . . ." *Id.* at 383. According to South Carolina lawyer Hugh Legaré, the rule of caveat emptor was desirable because it rejected the "refined equity" of the civil law in favor of "the policy of society." Though there was "something captivating in the equity of the principle, that a sound price implies a warranty of the soundness of the commodity," he was "certain that this rule is productive of great practical inconveniences . . . ." 2 WRITINGS OF HUGH SWINTON LEGARÉ 110 (M. Legaré ed. 1845). In South Carolina, he noted, "where we have had ample opportunity to witness its operation, there are very few experienced lawyers but would gladly expunge from our books the case which first introduced it here." *Id. See also* Barnard v. Yates, 1 N. & McC. 142, 146 (S.C. 1818) (noting "the perversion and abuse of [the] rule" which many "thought to have opened a door for endless litigation" in those cases where "the contracting parties had not placed themselves upon a perfect footing of equality in point of value").

were difficult to explain within a theory of exchange based on giving and receiving equivalents in value. Futures contracts for fungible commodities could only be understood in terms of a fluctuating conception of expected value radically different from the static notion that lay behind contracts for specific goods; a regime of markets and speculation was simply incompatible with a socially imposed standard of value. The rise of a modern law of contract, then, was an outgrowth of an essentially procommercial attack on the theory of objective value which lay at the foundation of the eighteenth century's equitable idea of contract.

We have seen, however, that there was a period during which vestiges of the eighteenth century conception of contract coexisted with the emerging will theory.[155] It was not until after 1820 that attacks on the equitable conception began to be generalized to include all aspects of contract law. If value is subjective, nineteenth century contracts theorists reasoned, the function of exchange is to maximize the conflicting and otherwise incommensurable desires of individuals. The role of contract law was not to assure the equity of agreements but simply to enforce only those willed transactions that parties to a contract believed to be to their mutual advantage. The result was a major tendency toward submerging the dominant equitable theory of contract in a conception of contractual obligation based exclusively on express bargains. In his *Essay on the Law of Contracts* (1822), for example, Daniel Chipman criticized the Vermont system of assigning customary values to goods that were used to pay contract debts. Only the market could establish a fair basis for exchange, Chipman urged. "[L]et money be the sole standard in making all contracts," for "[i]f, therefore, it were possible for courts in the administration of justice, to take this ideal high price as a standard of valuation, every consideration of policy, and a regard for the good of the people would forbid it." [156]

We will see that Nathan Dane's *Abridgment* (1823) and Joseph Story's *Equity Jurisprudence* (1836) also contributed to the demise of the old equitable conceptions. But nowhere were the underlying bases of contract law more brilliantly and systematically rethought than in Gulian C. Verplanck's *An Essay on the Doctrine of Contracts* (1825).

Verplanck was the first English or American writer to see in the "different parts of the system" of contract law "clashing and wholly incongruous" doctrines.[157] He emphasized "the singular incongruity" of a legal system that "obstinately refuses redress

[155] *See* note 18 & pp. 923, 924, 932–34 *supra*.

[156] D. CHIPMAN, *supra* note 68, at 109–11.

[157] G. VERPLANCK, *supra* note 53, at 57.

in so many, and such marked instances of unfairly obtained advantages" and yet "occasionally permit[s] contracts to be set aside upon the ground of inadequacy of price . . . ." [158] There were, he asserted, many "difficulties and contradictions" to be found in existing legal doctrine over "the question of the nature and degree of equality required in contracts of mutual interest," [159] as well as over the standards of "inadequacy of price" and "inequality of knowledge." "Where," he asked, "shall we draw the line of fair and unfair, of equal and unequal contracts?" [160]

Verplanck's *Essay* was written as an attack on the doctrine of caveat emptor, which had then only recently been adopted by the United States Supreme Court in *Laidlaw v. Organ* [161] (1817), one of the first cases to come before the Court involving a contract for future delivery of a commodity. The case, Verplanck wrote, raised "the important and difficult question of the nature and degree of equality in compensation, in skill or in knowledge, required between the parties to any contract . . . in order to make it valid in law, or just and right in private conscience." [162] He attacked caveat emptor on the ground that it should be fraudulent to withhold "any fact . . . necessarily and materially affecting the common estimate which fixes the present market value of the thing sold . . . ." [163]

In refusing to separate law and morals,[164] Verplanck was boldly independent of other theorists of the market economy.[165] But at its deepest level, Verplanck's *Essay* marks the triumph of a subjective theory of value in a market economy. Wishing to base legal doctrine on "the plainer truths of political economy," [166] he insisted that although just price doctrines bore "the impression of a high and pure morality," [167] they were "mixed with error"

[158] *Id.* at 199.

[159] *Id.* at 14 (emphasis deleted).

[160] *Id.* at 10.

[161] 15 U.S. (2 Wheat.) 178 (1817). The case grew out of a futures contract for sale of tobacco purchased by a merchant who had advance knowledge that the United States and England had signed a peace treaty ending the War of 1812. "The question in this case," Chief Justice Marshall wrote, "is, whether the intelligence of extrinsic circumstances, which might influence the price of the commodity, and which was exclusively within the knowledge of the vendee, ought to have been communicated by him to the vendor?" *Id.* at 195. The Chief Justice held that there was no duty to communicate the information, since "[i]t would be difficult to circumscribe the contrary doctrine within proper limits . . . ." *Id.*

[162] G. VERPLANCK, *supra* note 53, at 5.

[163] *Id.* at 125–26.

[164] Verplanck referred to the issue of fraud as "the [only] purely ethical part of the question . . . ." *Id.* at 117.

[165] *See, e.g.*, pp. 949–50 *infra* (N. Dane).

[166] G. VERPLANCK, *supra* note 53, at 106.

[167] *Id.* at 96.

and arose "from the introduction of a false metaphysic in relation to equality . . . ." [168] Thus, he disputed the view of "[l]awyers and divines . . . that all bargains are made under the idea of giving and receiving equivalents in value." [169] There could be no "such thing in the literal sense of the words, as adequacy of price [or] equality or inequality of compensation," since "from the very nature of the thing, price depends solely upon the agreement of the parties, being created by it alone. Mere inequality of price, or rather what appears so in the judgment of a third person, cannot, without reference to something else, be any objection to the validity of a sale, or of an agreement to sell." [170]

Verplanck's *Essay* represents an important stage in the process of adapting contract law to the realities of a market economy. Verplanck saw that if value is solely determined by the clash of subjective desire, there can be no objective measure of the fairness of a bargain. Since only "facts" are objective, fairness can never be measured in terms of substantive equality. The law can only assure that each party to a bargain is given "full knowledge of all material facts." [171] Significantly, Verplanck defined "material facts" so as not to include "peculiar advantages of skill, shrewdness, and experience, regarding which . . . no one has a right to call upon us to abandon. Here, justice permits us to use our superiority freely." [172] Thus, while he refused in theory to separate law and morality, Verplanck confined fraud to a range sufficiently narrow to permit the contract system to reinforce existing social and economic inequalities.

Though Verplanck's reconsideration of the philosophical foundations of contract law was by far the most penetrating among the American treatise writers, Nathan Dane and Joseph Story were more influential in contributing to the overthrow of an equitable conception of contract. In the very first chapter of his nine volume work, Dane elaborated some of the principles of contract law. One of his most important themes involved the "[d]ifference between morality and law." [173] He explained that while "in some special cases the *law of the land* and *morality* are the same," they

[168] *Id.* at 104.

[169] *Id.* at 8.

[170] *Id.* at 115. *See also id.* at 133.

[171] *Id.* at 225.

[172] *Id.* at 135.

> All know what a wide difference exists among men in these points, and whatever advantage may result from that inequality, is silently conceded in the very fact of making a bargain. It is a superiority on one side — an inferiority on the other, perhaps very great, but they are allowed. This must be so; the business of life could not go on were it otherwise.

Id. at 120.

[173] 1 N. DANE, *supra* note 147, at 100 (emphasis deleted).

differ in most cases, "when policy, or arbitrary rules must, also, be regarded." [174] "*'Virtue is alone the object of morality,'*" he continued, but "law has, . . . often, for its object, the peace of society, and what is practicable: Hence, though every . . . undue advantage in a bargain, to the hurt of another party, practised by one, is an act of injustice in the eyes of *morality*; yet it is not the mean [*sic*] of restitution in the eyes of the *law*; because [it is] often, impracticable *in every minute degree*." [175]

Dane also attacked all conceptions of a substantive theory of exchange. Equity decisions, Dane exclaimed, had become "trash" since they were "the productions of inferior lawyers" and "ignorant and indolent judges" who offered "no rule of property or conduct . . . ." [176] "Inadequate price in a bargain," he wrote, "does not defeat it, merely because inadequate . . . ." [177] But Dane remained willing to regard an unequal bargain as evidence that a "person did not understand the bargain he made, or was so oppressed, that he thought it best to make it . . . ." [178] Indeed, in his characteristic style, he continued to repeat the substance of the old learning while contributing to its overthrow. "[W]hen an agreement appears very unequal, and affords any ground to suspect any imposition, unfairness, or undue power or command, the courts will seize any very slight circumstances to avoid enforcing it." [179]

Dane was still reflecting an eighteenth century world view in which unequal bargaining power was conceived of as an illegitimate form of duress and in which lack of understanding was not yet identified only with mental disability. And yet in the world of speculation and futures markets, in which all value must simply turn on "an honest difference of opinion," [180] legal doctrine eventually renounced all claims to make judgments about oppression. With the publication of Joseph Story's *Equity Jurisprudence* (1836), American law finally yielded up the ancient notion that the substantive value of an exchange could provide an appropriate measure of the justice of a transaction. "Inadequacy of consideration," Story wrote, "is not then, of itself, a distinct principle of relief in Equity. The Common Law knows no such principle . . . . The value of a thing . . . must be in its nature fluctuating, and will depend upon ten thousand different circumstances . . . . If Courts of Equity were to unravel all these

[174] *Id.*

[175] *Id.*

[176] *Id.* at 107–08.

[177] *Id.* at 661.

[178] *Id.*

[179] *Id.*

[180] Seymour v. Delancey, 3 Cow. 445, 533 (N.Y. 1824).

transactions, they would throw every thing into confusion, and set afloat the contracts of mankind." [181]

The replacement of the equitable conception of contract with the will theory can be seen in Dane's assault on the eighteenth century practice of suing on a theory of implied contract where there had been a express agreement. In a long and unusually polemical technical discussion, Dane argued that once there is an express contract there could be no quantum meruit recovery off the contract on a theory of natural justice and equity.[182] Dane's attack on quantum meruit becomes comprehensible only as an effort to destroy an equitable conception of exchange in light of a newly emerging theory of value based on the subjective desires of contracting parties. Without a socially imposed standard of value, implied contracts make no sense. Where "there is no fixed or unchangeable comparative value between one price of property and another" and all value "depends on the wants and opinions of men," [183] it becomes impossible to measure damages by reference to customary value. The only basis for measuring contractual obligation, then, derives from the "will" of parties, and the crucial legal issue shifts to whether there has been a "meeting of minds."

The victory of the emerging will theory of contractual obligation was not at first complete. When Theron Metcalf delivered his lectures on contracts in 1828 he still reflected the tension between the old learning and the new.[184] Implied contracts, he wrote, were "inferred from the conduct, situation, or mutual relations of the parties, and enforced by the law on the ground of justice; to compel the performance of a legal and moral duty . . . ." [185] In support of this, he cited Chief Justice Marshall's statement that implied contracts "grow out of the acts of the parties. In such cases, the parties are supposed to have made those stipulations which, as honest, fair, and just men, they ought

[181] 1 J. STORY, COMMENTARIES ON EQUITY JURISPRUDENCE 249–50 (1836).

[182] 1 N. DANE, *supra* note 147, at 223–29.

[183] G. VERPLANCK, *supra* note 53, at 133.

[184] Metcalf's lectures were first published between 1839 and 1841, although they were first delivered in 1828 at a law school he had founded in Dedham, Massachusetts. *See* 1 U.S. L. INTELL. & REV. 142 (1829). When, in 1867, Metcalf published his *Principles of the Law of Contracts*, he acknowledged that "[t]he first manuscript of the . . . work was prepared, in the years 1827 and 1828" and was published in *American Jurist* between 1839 and 1841. T. METCALF, PRINCIPLES OF THE LAW OF CONTRACTS iii (1867) [hereinafter cited as LAW OF CONTRACTS]. "That publication," he wrote, "has recently been revised and enlarged by reference to reports and treatises published since 1828; but no change has been made in the original arrangement." *Id.*

[185] LAW OF CONTRACTS 4; 20 AM. JUR. 5 (1838).

to have made." [186] Though both Metcalf and Marshall were beginning to pretend that contractual obligation derives only from the will of the parties, their predominant form of expression continued to recognize standards of justice external to the parties. Indeed, Metcalf still maintained that "[i]n sound sense, divested of fiction and technicality, the only true ground, on which an action upon what is called an implied contract can be maintained, is that of justice, duty, and legal obligation." [187]

By the time William W. Story's *Treatise on the Law of Contracts* appeared in 1844, however, the tension between the two theories had dissolved. "Every contract," he wrote, "is founded upon the mutual agreement of the parties . . . ." [188] Both express and implied contracts were "equally founded upon the actual agreement of the parties, and the only distinction between them is in regard to the mode of proof, and belongs to the law of evidence." [189] For implied contracts, he concluded, "the law only supplies that which, although not stated, must be presumed to have been the agreement intended by the parties." [190] Since the only basis for the contractual obligation was the will of the parties, Story now maintained, implied promises "only supply omissions, and do not alter express stipulations"; he was thus prepared to announce the "general rule" that there could be no implied contract where an express agreement already existed.[191]

With Story's announcement of the "general rule," the victory of the will theory of contractual obligation was complete. The entire conceptual apparatus of modern contract doctrine — rules dealing with offer and acceptance, the evidentiary function of consideration, and canons of construction and interpretation — arose to articulate the will theory with which American doctrinal writers expressed the ideology of a market economy in the early nineteenth century.

[186] Ogden v. Saunders, 25 U.S. (12 Wheat.) 341 (1827); *see* LAW OF CONTRACTS 4 n.(b); 20 AM. JUR. at 5 n.1.

[187] LAW OF CONTRACTS 5–6. This passage does not appear in *American Jurist*, though Metcalf did write that "it is manifestly only by a fiction, that a contract or promise is implied. And, indeed, the whole doctrine of implied contracts, in all their varieties, seems to be merely artificial and imaginary." 20 AM. JUR. at 9.

[188] W. STORY, *supra* note 53, at 4.

[189] *Id.*

[190] *Id. See also* 2 S. GREENLEAF, A TREATISE ON THE LAW OF EVIDENCE 87 (1850) ("The distinction between general or *implied contracts* and special or *express contracts* lies not in the nature of the undertaking, but in the mode of proof").

[191] W. STORY, *supra* note 53, at 6. *Cf.* pp. 933–34 *supra*.

C. The Application of the Will Theory of Contract to Labor Contracts

Thus far, we have seen the changes in contract law which were necessary to meet the needs of the newly emerging market economies in England and America. There is evidence, however, that the change from the eighteenth to the nineteenth century also involved a pervasive shift in the sympathies of the courts. In the eighteenth century the subjection of individual bargains to the extensive supervisory powers of courts and juries expressed the legal and ethical culture of the small town, of the farmer, and of the small trader. In the nineteenth century, the will theory of contract was part of a more general process whereby courts came to reflect commercial interests. The changing alliances are painfully obvious in nineteenth century courts' discriminatory application of the recently discovered chasm between express and implied contracts.

The most important class of cases to which this distinction applied was labor contracts in which the employee had agreed to work for a period of time — often a year — for wages that he would receive at the end of his term. If he left his employment before the end of the term, jurists reasoned, the employee could receive nothing for the labor he had already expended. The contract, they maintained, was an "entire" one, and therefore it could not be conceived of as a series of smaller agreements. Since the breach of any part was therefore a breach of the whole, there was no basis for allowing the employee to recover "on the contract." Finally, citing the new orthodoxy proclaimed by the treatise writers, judges were led to pronounce the inevitable result: where there was an express agreement between the parties, it would be an act of usurpation to "rewrite" the contract and allow the employee to recover in quantum meruit for the "reasonable" value of his labor.[192]

Courts in fact seemed driven to resolve all ambiguity in contracts in favor of the employer's contention that they were "entire." It made no "difference . . . whether the wages are estimated at a gross sum, or are to be calculated according to a certain rate per week or month, or are payable at certain stipulated times, provided the servant agree for a definite and whole term . . . ." [193] Under these circumstances, it should be emphasized, the assumption that the agreement was "for a definite and whole term" was simply a judicial construction not required by the terms of the agreement. Moreover, it did not "make any

[192] *See* Annot., 19 AM. DEC. 268, 272 (1880).

[193] 1 T. PARSONS, THE LAW OF CONTRACTS 522 n.(1) (1853).

difference, that the plaintiff ceased laboring for his employer, under the belief that, according to the legal method of computing time, under similar contracts, he had continued laboring as long as could be required of him." [194] Nor did it matter that the "employer, during the term, has from time to time made payments to the plaintiff for his labor." [195] The result of the cases was that any employee not shrewd or independent enough to demand immediate payment for his work risked losing everything if he should leave before the end of the contract period. The employer, in turn, had every inducement to create conditions near the end of the term that would encourage the laborer to quit.

The disposition of courts ruthlessly to follow conceptualism in the labor cases was not, however, quite matched in cases involving building contracts. Building contracts are similar to labor agreements in that there is no way of restoring the status quo after partial performance. Nevertheless, nineteenth century courts allowed builders to recover "off the contract" when they had committed some breach of their express obligation. The leading case is *Hayward v. Leonard* [196] (1828), in which the Supreme Judicial Court of Massachusetts held that a builder could recover in quantum meruit "where the contract is performed, but, without intention, some of the particulars of the contract are deviated from." [197] If there was "an honest intention to go by the contract, and a substantive execution of it," [198] the court held, it would not decree a forfeiture. It should be noted that the Massachusetts court in *Hayward v. Leonard* expressly rejected Dane's view that the existence of an express contract barred recovery in quantum meruit. There was, Chief Justice Parker declared, "a great array of authorities on both sides, from which it appears very clearly that different judges and different courts have held different doctrines, and sometimes the same court at different times." [199] The result was that in Massachusetts and in most other states two separate lines of cases were developed, one dealing with service contracts, for which recovery in quantum meruit was barred, and another applying to building contracts, for which recovery "off the contract" of the reasonable value of the performance was permitted.

Few courts attempted to rationalize what Theophilus Parsons was later to call these "very conflicting" decisions.[200] The leading

[194] *Id.*

[195] *Id.*

[196] 24 Mass. (7 Pick.) 181 (1828), *annotated*, 19 AM. DEC. 268 (1880).

[197] 24 Mass. (7 Pick.) at 186.

[198] *Id.* at 187.

[199] *Id.* at 184.

[200] 2 T. PARSONS, *supra* note 193, at 35 & n.(d). There were two exceptions in this trend. The New York courts applied the express contract theory to building

explanation came from *Hayward v. Leonard* itself. In the labor cases the employee usually broke his contract "voluntarily" and "without fault" of his employer. Breach of building contracts was often "without intention" and compatible with an "honest intention" to fulfill the contract.[201] Thus, it was not that courts had abandoned an underlying moral conception of contracts, but that the morality had fundamentally changed. The focus had shifted from an emphasis on the role of quantum meruit in preventing "unjust enrichment." The express contract had become paramount; denial of quantum meruit recovery was now employed to enforce the contract system. It was now regarded as just for the employer to retain the unpaid benefits of his employee's labor as a deterrent to voluntary breach of contract. But it was still unjust for the beneficiary of a building contract to enrich himself because of an honest mistake in performing the contract.[202]

While the judges who adhered to the distinction between labor and building contracts never acknowledged an economic or social policy behind the distinction, it seems to be an important example of class bias. A penal conception of contractual obligation could have deterred economic growth by limiting investment in high risk enterprise. Just as the building trade was beginning to require major capital investment during the second quarter of the nineteenth century, courts were prepared to bestow upon it that special solicitude which American courts have reserved for infant industry. Penal provisions in labor contracts, by contrast, have only redistributional consequences, since they can hardly be expected to deter the laboring classes from selling their services in a subsistence economy.

Although nineteenth century courts and doctrinal writers did not succeed in entirely destroying the ancient connection between contracts and natural justice, they were able to elaborate a system that allowed judges to pick and choose among those groups in

as well as to labor contracts. Smith v. Brady, 17 N.Y. 173, 187 (1858). The second exception is the solitary challenge in New Hampshire to the doctrine against quantum meruit recovery in labor cases. *See* Britton v. Turner, 6 N.H. 481 (1834).

[201] 24 Mass. (7 Pick.) at 185.

[202] Even when courts modified in building contracts cases the dominant view of the treatise writers that express contracts barred all recovery on an implied contract, they shared at a deeper level the treatise writers' basic assumption about the relationship between express and implied agreements. The contract price, all agreed, set the limit on recovery in quantum meruit. *See, e.g.*, Hayward v. Leonard, 24 Mass. (7 Pick.) 181, 187 (1828). Similarly, in the great case of Britton v. Turner, 6 N.H. 481 (1834), where New Hampshire Chief Justice Joel Parker stood almost alone in resisting the orthodox view barring quantum meruit recovery on labor contracts, he permitted the employer to deduct from recovery "any damage which has been sustained by reason of the nonfulfillment of the contract." *Id.* at 494.

the population that would be its beneficiaries. And, above all, they succeeded in creating a great intellectual divide between a system of formal rules — which they managed to identify exclusively with the "rule of law" — and those ancient precepts of morality and equity, which they were able to render suspect as subversive of "the rule of law" itself.

Notas críticas sobre el fundamento de la fuerza obligatoria del contrato. Fuentes e interpretación del artículo 1545 del *Código civil* chileno

Carlos PIZARRO WILSON[*]

Aunque es frecuente indicar que el fundamento de la fuerza obligatoria del contrato radica en el dogma de la autonomía de la voluntad, una re-construcción dogmática histórica permite demostrar que este principio angular de la contratación solo fue forjado a finales del siglo XIX producto de una errónea interpretación de la filosofía kantiana. Además, la evolución de la contratación permite constatar el declive del rol de la voluntad en la teoría general de los contratos. La exposición del análisis tradicional de la fuerza obligatoria del contrato permite examinar las críticas que el dogma de la autonomía de la voluntad suscita.

1. **La interpretación tradicional del artículo 1545 del Código civil.** La fuente de inspiración de la regla prevista en el artículo 1545 del Código civil es el artículo 1134 del *Code civil*. Ambos preceptos consagran la fuerza obligatoria del contrato en términos similares. La doctrina francesa elaboró la doctrina de la autonomía de la voluntad, cuyo reconocimiento estaría plasmado en el citado precepto del *Code*. Esta interpretación doctrinal será acogida por la doctrina nacional, la cual reconoce en el artículo 1545 del *Código civil* la expresión del principio de la autonomía de la voluntad. En suma, la interpretación tradicional del artículo 1545 es el resultado de la influencia de los autores franceses de finales del siglo XIX y los primerios decenios del siglo XX. Sin embargo, el dogma de la autonomía de la voluntad aparece desmentido por un análisis de las fuentes del artículo 1134 del *Code*. La doctrina francesa que influyó en la civilística nacional y que determinó la interpretación actual del artículo 1545 del *Código civil* aparece desmentida por un análisis de las fuentes y de la doctrina contemporánea. Esta es la razón por la cual parece necesario referirse a la doctrina francesa que inspiró la interpretación errónea del citado artículo 1545 del *Código civil*.

2. En la dogmática francesa existen importantes estudios consagrados al fundamento obligatorio del contrato.[1] Por el contrario, la doctrina chilena ha mostrado un interés precario sobre este problema.[2] La fuerza vinculante del contrato sería una consecuencia ineluctable de la voluntad de las partes contratantes. La doctrina individualista, inherente al derecho de las obligaciones, concibe la fuerza obligatoria del contrato como la principal expresión de la autonomía de la voluntad. Sin embargo, este *axioma* del derecho civil chileno es difícil de justificar. El análisis de las fuentes del citado artículo 1545 desmiente tal afirmación. La hipótesis según la cual la autonomía de la voluntad sería un principio fundamental de la codificación que explica la fuerza vinculante del contrato es el resultado de una hermenéutica equivocada de la doctrina francesa a fines del siglo XIX. La filosofía kantiana y el liberalismo económico, a la moda en esa época, influyeron de manera importante en la doctrina francesa que otorgó a la autonomía de la voluntad el rango de fundamento único del derecho de los contratos.[3] Una revisión de las fuentes del artículo 1134 del *Code civil français* permite corroborar esta afirmación.

3. **Domat y la fuerza obligatoria del contrato.** Para nadie es un misterio que la regla prevista en el artículo 1134 del *Code* que consagra la fuerza obligatoria del contrato se inspira en la pluma de DOMAT.[4] Sin embargo, al estudiar su obra, es fácil darse cuenta que no reconoce a la autonomía de la voluntad el valor de fundamento de la fuerza obligatoria del contrato. Tampoco otorga la categoría

[*]Profesor de derecho civil en la universidad Diego Portales y de Chile. Doctor en derecho por la universidad Paris II (Panthéon-Assas).

[1] GOUNOT, E., *Le principe de l'autonomie de la volonté en droit privé ; contribution à l'étude critique de l'individualisme,* thèse, Dijon, 1912 ; RIEG, A., *Le rôle de la volonté dans l'acte juridique en droit civil français et allemand*, thèse, Paris, LGDJ, 1961 ; ROUHETTE, G., *Contribution à l'étude critique de la notion de contrat*, thèse, dactyl., Paris, 1965 ; ANCEL, P., « Force obligatoire et contenu obligationnel du contrat », en *RTD civ.*, 1999, p. 777.

[2] Cfr. LÓPEZ SANTA-MARÍA, J., *Los contratos. Parte general*, Santiago, Jurídica, 2001[3], t. I, nº 45, p. 265.

[3] GOUNOT, E., (n. 1), p. 26 y ss. ; ROUHETTE, G., (n. 1), p. 417 y ss. ; BACACHE-GIBEILI, M., *La relativité des conventions et lers groupes de contrats*, Paris, LGDJ, préface Yves Lequette, p. 231 y ss. ; TERRE, F., « Sur la sociologie juridique du contrat », en *Arch. Ph. du droit*, 1968, en particular p. 78 ; BATIFFOL, H., « La crise du contrat et sa portée », en *Arch. Ph. du droit*, 1968, p. 21.

[4] Art. 1134 Les conventions légalement formées tiennent lieu de loi à ceux qui les ont faites. DOMAT, J., *Les lois civiles dans leur ordre naturel. Ouvres complètes*, Paris, Alex-Gobelet, 1835, Liv. I, II, § VII,"Les conventions étant formées, tout ce qui a été convenu tient lieu de loi à ceux qui les ont faites".

de principio a la autonomía de la voluntad. Este autor, que no era partidario de un individualismo liberal, recuerda en varios pasajes de su obra la equidad y la moral para explicar el deber de ejecución de los contratos. Para DOMAT, jansenista convencido, la razón del carácter obligatorio del contrato no radica en la voluntad de las partes, sino en considerar que el contrato es justo.[5] Según afirma una doctrina autorizada, el principio de la autonomía de la voluntad es extraño a la obra de Domat y Pothier.[6] Ghestin afirma que para los inspiradores de la teoría general del contrato: "la fuerza obligatoria del contrato es exterior al hombre. Ella es esencialmente moral y reposa no sobre la libertad individual, concebida como un derecho natural, sino sobre un deber de conciencia: *el respeto de la palabra dada*. Para ellos, las convenciones, además *del respeto de las leyes y de las buenas costumbres*, deben inspirarse en la *buena fe* y en la *equidad*, y en un sentido más general en *la ley divina*."[7] Para aclarar el pensamiento de DOMAT, debemos tener presente qué entiende por ley, palabra clave en la definición de la fuerza obligatoria del contrato. Al utilizar la palabra *loi* en el parágrafo que servirá de inspiración al artículo 1134, DOMAT le atribuye la significación de lo que es justo.[8] El siguiente parágrafo es elocuente: *"On entend communément par ces mots de lois et de règles, ce qui es juste, ce qui est ordonné, ce qui est réglé"*.[9] En suma, cuando Domat utiliza la expresión *loi*, evoca la idea de justicia y equidad y no se refiere a la fuerza coercitiva de la ley, confusión que habría aparecido en los teólogos medievales.[10] El contrato constituye una ley para las partes en el sentido que permite al juez atribuir a cada contratante lo que le corresponde en su justa medida.[11] En suma, el análisis de la obra de Domat permite afirmar con certeza que la autonomía de la voluntad constituye un dogma ignorado en el principal inspirador de la regla establecida en el artículo 1134 del *Code*. No podía ser de otra forma, pues, según veremos la expresión "autonomía de la voluntad" aparece tardíamente en la lengua francesa.

4. **Los redactores del *Code* y la autonomía de la voluntad.** Los redactores del *Code* no se apartan del fundamento de la fuerza obligatoria de los contratos propuesto por DOMAT. Para estos también se justifica el efecto obligatorio del contrato en la moral y en la equidad. PORTALIS señala que las convenciones deben someterse a la justicia objetiva que precede su celebración y de la cual arranca su principal fuerza.[12] Los trabajos preparatorios del Code aluden a "*la foi donnée*" y a "*l'équité naturelle*".[13] Lejos de pensar en un hombre racional con una voluntad libre, los redactores manifiestan una concepción pesimista del hombre, entendido como un sujeto atrapado por sus pasiones, cuya voluntad voluble está sujeta a sus incliniciones y debilidades. Así, por ejemplo TREILHARD expresa que "*le législateur ne saurait ignorer que les passions ont trop souvent étouffé la raison et fait taire la bonté*".[14] Por su parte, el tribuno FAVARD afirma que "*il est du devoir du législateur de forcer les hommes à voir*

[5] Así podemos citar el siguiente párrafo del *Traité des lois:* "Y como él (Dios) marca en cada compromiso lo que prescribe a aquellos que vincula, se reconoce en los caracteres de diferentes tipos de compromisos, los fundamentos de diversas reglas de aquello que *la justicia y la equidad* demandan de cada persona", DOMAT, J., *Traité des lois, Ouvres complètes*, Paris, Alex-Gobelet, 1835, chapitre I, II, § 3. Véase BONASSIES, P., *Le dol dans la conclusion des contrats*, thèse, dactyl., Lille, 1955; CHAZAL, J. P., « De la signification du mot loi dans l'article 1134 du Code civil », en *RTD civ.*, 2001, p. 265, en particular p. 267 *in fine*. Esta autora afirma de manera categórica que "hay que terminar con el anacronismo que consiste en creer que asimilando la convención a la ley, Domat y, luego, los redactores del código civil habrían querido consagrar el principio de la libertad contractual o de la autonomía de la voluntad".

[6] GHESTIN, J., *Traité de droit civil. La formation du contrat*, Paris, LGDJ, 1993, nº 58, p. 40.

[7] GHESTIN, J., "Jean Domat et le Code civil français", en *Mélange Sacco*, 1993. Los destacados son del autor. Véase ROUHETTE, G., (n. 1), t. I, nº 195, p. 594.

[8] *Ibidem*

[9] *Les lois civiles...*, (n.4), Liv. Prél., I, I. Des diverses sortes de règles, et de leur nature. "Entendemos comúnmente por estas palabras de leyes y reglas, lo que es justo, lo que es ordenado, lo que ha sido reglado".

[10] En el mismo sentido Pothier, siendo más proclive al voluntarismo contractual, señala que "La equidad debe reinar en las convenciones", *Traité des obligations*, nº 33. Sobre el particular, CHAZAL, J. P., (n. 5), p. 271.

[11] En efecto, se trata de la equidad en el sentido de la igualdad aristotélica. Para CHAZAL, la palabra *jus* deriva del vocablo *justitia* que en latín es sinónimo de equidad. Así aparecería en el Digesto, según lo explica Ulpiano (D. 1, 1, 1) y según nos da noticia Paulo (D. 1, 1, 11) *id quod semper aequum ac bonum est, jus dicitur*. Cit. por CHAZAL, (n. 5), p. 274, nota (62). Sin embargo, según da cuenta Guzmán Brito, la etimología de la palabra *Ius* es desconocida. Por su parte, la expresión *Iustitia*, según la define Ulpiano es "una constante y perpetua voluntad de dar a cada uno su derecho" *(constans et perpetua voluntas ius suum cuique tribuendi*), D. 1. 1. 10. Véase GUZMÁN BRITO, A., *Derecho privado romano*, Santiago, Jurídica, 1996, t. I, pp. 83102; también, ARISTOTE, *Etique de nicomaque*,traducción y prefacio de J. Voilquin, Paris, 1992, liv. V, chap. IV, 3: "le juste dans les contrats consiste en une certaine égalité" (lo justo en los contratos consiste en una cierta igualdad). Sobre la contradicción moderna entre la ley y la equidad, PERELMAN, C., "Cinq leçons sur la justice", in *Morale et Philosophie*, Paris, 1968, p. 18.

[12] "A Dieu ne plaise que nous voulions affaiblir le respect qui est dû à la foi des contrats! Mais il est des règles de justice qui sont antérieures aux contratas mêmes, et desquelles les contrats tirent leur principal force. Les idées de juste et del'injuste ne sont pas l'unique résultat des conventions humaines. Elles ont précédé ces conventions, et elles doivent en diriger les pactes", en Fenet, t. 14, p. 130, cit. por ROUHETTE, G., "La force obligatoire du contrat", en *Le contrat aujourd'hui: comparaisons franco-anglaises*, sous la direction de Denis TALLON et Donald HARRIS, Paris, LGDJ, 1987, p. 44.

[13] *Ibidem.*

[14] *Présentation au Corps législatif*, cit. por ROUHETTE, (n. 12), p. 45. « el legislador no podría ignorar que las

des lois dans les contrats"[15] El texto del *Code* tampoco hace referencia a la autonomía, salvo por excepción, al referirse a la formación del usufructo[16] o a propósito de la *resiliation* del contrato de obra.[17] Tampoco existe una amplia libertad para la voluntad de las partes, según suele afirmarse. Por el contrario, la esfera de libertad se encuentra restringida por leyes imperativas: el orden público, las buenas costumbres y otras cláusulas generales. Incluso la protección de la voluntad, tratándose de los vicios del consentimiento, se realiza de manera restrictiva. Solo por excepción, por ejemplo, el error constituye un vicio del consentimiento y en variadas hipótesis el error no resta validez al contrato.[18] El voluntarismo es ineficaz para explicar los contratos solemnes, cuya fuerza obligatoria es posterior a la formación del consentimiento. La concepción objetiva de la lesión, en desmedro de la voluntad de las partes, que posibilita la intervención del juez, dista bastante del voluntarismo. En definitiva, los redactores del *Code* y el texto mismo no consagran una teoría de la autonomía de la voluntad, siendo ésta ajena a la codificación. En términos generales, puede señalarse que los redactores del *Code*, lejos de inspirarse en el supuesto dogma de la autonomía de la voluntad, hacen referencia a la moral y a la equidad para justificar la fuerza obligatoria del contrato.[19] Sin embargo, una vez en vigencia el *Code Napoléon* debemos revisar el pensamiento de los comentaristas, a fin de trazar el surgimiento del dogma de la autonomía de la voluntad en la doctrina francesa.

5. **Los comentaristas del *Code*.** Para reflejar el pensamiento de los primeros comentaristas del *Code* podemos citar a Toullier, quien de manera categórica señala: "Toda obligación viene de la ley...es la ley que la produce...Así, toda obligación supone una ley anterior; toda obligación viene de la ley, ya sea inmediatamente por un simple acto de voluntad del legislador, ya sea por medio de la voluntad o el hecho del hombre. Las convenciones, ellas mismas sólo obligan en virtud de la ley que exige mantener la palabra dada. El legislador les confiere la autoridad de la ley, como lo dice enérgicamente el artículo 1134"[20]Los primeros comentaristas del *Code*,[21] siguen la doctrina de Duranton, quien afirma de manera elocuente que "las obligaciones nacen, o de un hecho del hombre, al cual la ley otorga la fuerza obligatoria, o de la sola ley".[22] Para Demolombe todas las obligaciones, ya sean legales o convencionales extraen su obligatoriedad de la ley civil.[23] Acollas, por su parte, afirma que es la ley el fundamento mediato de la fuerza obligatoria del contrato.[24] La opinión de los comentaristas no podía ser distinta, pues según nos informa BÜRGE, la expresión *autonomie de la volonté*, en ese entonces era desconocida en la lengua francesa. Solo alrededor de los años 30 del siglo XIX la idea de voluntad será vinculada al contrato.[25] Sin duda fueron los internacionalistas los primeros en utilizar la expresión *autonomie de la volonté.*[26] No existen dudas que la autonomía de la voluntad solo se consolida a fines del siglo XIX en la doctrina francesa, para luego, transformarse en un *bien común de la cultura jurídica francesa*.[27] El análisis de las fuentes del artículo 1134 del *Code* y las opiniones de los comentaristas permiten constatar que la doctrina de la autonomía de la voluntad

pasiones de manera frecuente han asfixiado la razón y hecho callar la bondad".

[15] FENET, t. 13, p. 336, cit. por ROUHETTE, (n. 12), p. 46."*es deber del legislador forzar los hombres a ver leyes en los contratos*".

[16] Art. 579 L'usufruit est établi par la loi, ou par la volonté de l'homme.

[17] Le maître peut résilier, par sa simple volonté, le marché à forfait, quoique l'ouvrage soit déjà commencé, en dédommageant l'entrepreneur de toutes ses depenses, de tous ses travaux, et de tout ce qu'il aurait pu gagner dans cette entreprise.

[18] Art. 1110 L'erreur n'est une cause de nullité de la convention que lorsqu'elle tombe sur la substance même de la chose qui en est l'objet.
Elle n'est point une cause de nullité, lorsqu'elle ne tombe que sur la personne avec laquelle on a intention de contracter, à moins que la considération de cette personne ne soit la cause principale de la convention.

[19] GHESTIN, J., *Traité*, (n. 6), p. 41.

[20] TOULLIER, *Le droit civil français*, Paris, 1830/43, 5 éd., t. 1-15, t. 6, p. 3 y ss., nº 3 y ss.

[21] LAROMBIERE, *Théorie et pratiques des obligations*, t. V, 1857, p. 545.

[22] *Cours de droit français suivant le code civil*, 4 éd., 1844, t. 10, nº 20, p. 10.

[23] DEMOLOMBE, *Cours de Code Napoléon*, t. 31, nº 15, p. 16.

[24] ACOLLAS, *Les contrats et les obligations contractuelles. Le droit mis à la portée de tout le monde*, 1885, p. 13.

[25] Según BÜRGE, el primer autor francés que vincula la autonomía de la voluntad y el contrato es Schutzenberger: « y la autonomía individual sanciona las convenciones formadas por el consentimiento libre y reflexivo de las partes contratantes, en cuanto su objeto no esté en contradicción con la moral », *Etudes de droit public*, Paris-Strasbourg, 1837, p. 226. Cit. por BÜRGE, A., « Le code civil et son évolution vers un droit imprégné d'individualisme libéral », en *RTD civ.*, 2000, p. 1.

[26] RANOUIL, V, *L'autonomie de la volonté: naissance et évolution d'un concept*, Paris, PUF, 1980, p. 21.En palabras de Rouhette, el principal aporte de Ranouil consiste precisamente en haber demostrado que la expresión "autonomía de la voluntad" aparece en Francia con la obra de Weiss. La razón consiste en la ausencia de textos o Código en el derecho internacional privado, lo que hace más necesario un principio justificativo. De otra parte, si bien es usual imputar a la doctrina alemana la idea de autonomía, la cual se refleja en dos conceptos distintos: *Parteiautonomie* en el ámbito del derecho internacional privado y *Privatautonomie* en derecho interno, ambas nociones no se refieren a la automnomía de la voluntad, creación de la doctrina jurídica francesa en el último cuarto del siglo XIX. Vid. ROUHETTE, G., (n. 12), nº 12, p. 38.

no era el fundamento de la fuerza obligatoria del contrato. Esta doctrina será creada por la doctrina francesa hacia finales del siglo XIX. Cabe preguntarse, entonces, cuáles son los meandros del surgimiento de esta teoría que en definitiva justificará para la doctrina francesa la fuerza obligatoria del contrato y la teoría general del contrato.

6. **Surgimiento y consolidación de la doctrina de la autonomía de la voluntad.** No cabe duda que la autonomía de la voluntad constituye una noción fundamental en la filosofía kantiana.[28] De ahí que la tentación en la disciplina del derecho haya sido a emparentar dicha noción con la doctrina jurídica de la autonomía de la voluntad.[29] Sin embargo, Ranouil nos entrega algunas razones para desconocer el vínculo entre la filosofía kantiana y el dogma de la autonomía de la voluntad como fundamento de la fuerza obligatoria del contrato. En primer lugar, sorprende que los primeros juristas que aluden a la autonomía de la voluntad (Weiss y Brocher) no hayan realizado ninguna referencia a Kant.[30] En segundo lugar, en el evento que los autores citados hubieran leído a Kant, se trataría de su obra jurídica, en la cual la expresión autonomía de la voluntad esta ausente.[31] En efecto, la expresión *autonomía de la voluntad* aparece explicitada en la *Critica de la razón práctica*.[32] Sin embargo, el argumento principal para descartar el origen kantiano del dogma de la autonomía de la voluntad como fundamento de la fuerza obligatoria del contrato radica en la distancia conceptual entre esta noción utilizada por Kant y aquella propuesta por la doctrina jurídica. Para Kant, la autonomía de la voluntad tiene una significación ética. Por esto afirma que "la legislación que hace de una acción un deber y de ese deber, a la vez, un móvil, es *ética*. Pero la que no incluye al último en la ley y, por tanto, admite también otro móvil distinto de la idea misma del deber, es *jurídica*". Luego agrega que "A la mera concordancia o discrepancia de una acción con la ley, sin tener en cuenta los móviles de la misma, se le llama *legalidad* (conformidad con la ley), pero a aquélla en la que la idea del deber según la ley es a la vez el móvil de la acción, se le llama la *moralidad* (eticidad) de la misma".[33] Para Kant, entonces, la coincidencia entre el deber y el móvil constituye una ley ética. La legalidad, en cambio, puede justificar el deber en un móvil distinto al arbitrio o voluntad del sujeto, tiene que ser "una legislación que coaccione".[34] Esta es la razón por la cual, según Kant, "la ética manda que yo cumpla el compromiso contraído en un contrato, aunque la otra parte no pudiera acto seguido obligarme a ello: pero toma de la doctrina del derecho, como dados, la ley (*pacta sum servanda*) y el deber correspondiente a ella. Por tanto, la legislación de que las promesas aceptadas han de cumplirse no reside en la ética, sino en el *Ius*".[35] Todavía podemos citar la reflexión de Kant relativa al contrato de donación: "Este contrato (*donatio*), por el que *enajeno* lo mío, mi cosa (o mi derecho) *gratuitamente* (gratis), contiene una relación de mí, el donante (*donans*), con otro, el donatario (*donatarius*), según el *derecho privado*; relación por lo que (lo) mío pasa a él por su aceptación (*donum*).- Pero no se puede suponer que con esto piense que estoy obligado por ello a cumplir mi promesa y, por tanto, también a regalar gratuitamente mi libertad y, por así decirlo, a venderme a mí mismo (*nemo suum iactare praesumitur*), lo cual, sin embargo, sucedería en el estado civil, conforme al derecho; porque en él el donatario puede *forzarme* a la prestación de la promesa".[36] En suma, para Kant la fuerza obligatoria del contrato se encuentra en el *Ius* (doctrina del derecho) y no en la pretendida autonomía de la voluntad. Esta última constituye para Kant un principio de las leyes morales, mas no de la legislación o de los contratos. De ahí que señalar la filosofía kantiana como el origen de la doctrina jurídica de la autonomía de la voluntad no deje de ser una afirmación en contradicción con la historia de la ideas,

[27] BÜRGE, A., (n. 25), p. 11; CHAZAL, J. P., (n. 5), p. 267.
[28] KANT, E., *Critique de la raison pratique*, traducción de François PICAVET, Paris, PUF, 1971, teorema IV, p. 33.
[29] GOUNOT, E, (n. 1), p. 53 y ss.
[30] RANOUIL, V., (n. 26), pp. 53-56; ROUHETTE, G., (n. 12), p. 39.
[31] ROUHETTE, G., *ibidem*.
[32] KANT, E., (n. 28), p. 33.
[33] KANT, I., *La metafísica de las Costumbres*, traducción y notas de Adela CORTINA ORTS y Jesús CONILL SANCHO, Madrid, Tecnos, 1994[2], p. 24.
[34] *Ibidem.*
[35] *Op. cit.*, p. 25. Para Kant, la expresión *Ius* corresponde a la doctrina del derecho, al conjunto de leyes, para las que es posible una legislación exterior. *Idem*, p. 229.
[36] *Op. cit.*, p. 298.

producto de su vulgarización en la enseñanaza de la filosofía en Francia.[37] Sin embargo, Fouillé en su obra *L'idée moderne du droit*, reconoce en la obra de Kant a la autonomía de la voluntad el carácter de fundamento del derecho positivo.[38] Este fue, al parecer, el primer paso, algo confuso y vago, para configurar el desliz desde la concepción kantiana de la autonomía de la voluntad hacia el dogma jurídico que servirá de fundamento a la fuerza obligatoria del contrato, desconociendo la distinción fundamental entre la regla moral y la regla jurídica. En la civilística francesa, Gény fue el primero en utilizar la expresión "autonomía de la voluntad" en la primera edición de su monumental *Méthode d'interprétation et sources en droit privé positif* (1899) al referirse a ciertos problemas jurídicos a través de los cuales ilustra la proposición de método.[39] Empero, será la tesis de Gounot en 1912 la primera sistematización del seudo principio de la autonomía de la voluntad en el derecho francés. Constituye una paradoja que precisamente Gounot haya realizado su trabajo con el único objeto de criticar el dogma de la autonomía de la voluntad.[40] Una vez que Gounot estructura la autonomía de la voluntad con sus diversas manifestaciones jurídicas, aparejado con la aparición de la contratación en masa (contrats d'adhésion), comienza el declive del principio. La doctrina francesa preocupada de proteger al contratante débil comienza a buscar los distintos mecanismos jurídicos en la teoría general del contrato para establecer los límites necesarios al principio de la autonomía de la voluntad. Actualmente, la mayor parte de la doctrina recoge la sistematización de Gounot.[41] Aunque luego de la espléndida tesis de Rouhette, la afirmación que el *Code* civil recogería en materia contractual la teoría de la autonomía de la voluntad aparece bastante atenuada.[42]

[37] Véase ROUHETTE, G., (n. 12), pp. 40-41, notas 128 y ss.
[38] FOUILLE, A., *L'idée moderne du droit*, 1883[2], pp. 25-29, cit. por ROUHETTE, G., (n. 12), p. 41.
[39] GENY, F., *Méthode d'interprétation et sources en droit privé positif*, Paris, LGDJ, 1954, en particular nº 170 a 173. Véase RANOUIL, (n. 26), p. 93.
[40] GOUNOT, E., (n. 1) ; TISON, R., *Le principe de l'autonomie de la volonté en droit privé français, thèse*, Paris, 1931; RANOUIL, (n. 26).
[41] LARROUMET, Ch.,*Droit civil. Les obligations. Le contrat*, Paris, Economica, 1998[4], nº 111, p. 94 y ss.; para una ilustración de la posición tradicional, MAZEAUD, L. Y H., CHABAS, F., *Leçons de droit civil. Obligations. Théorie générale*, 1998[9], nº 28, p. 22: "C'est le principe, essentiel pour le rédacteurs du Code civil, de *l'autonomie de la volonté*".
[42] Véase TERRE, F., SIMLER, Ph. et LEQUETTE, Y., *Droit civil. Les obligations*, 8e éd., 2002,. n° 27, p. 32, quienes afirman que la teoría de la autonomía de la voluntad es la obra de una parte de la doctrina de finales del último siglo (XIX) y del inicio de este siglo (XX), la cual se proponía criticar la concepción clásica, nº 19, p. 29; STARCK, B., ROLAND, H. et BOYER, L., *Droit civil. Les obligations. 2. Contrat*, 6e éd., 1998. n° 18, p. 7, quienes afirman que "La volonté aménage ces rapports d'interdépendence nécessaire, et elle ne peut les aménager que dans le cadre de la loi et dans la mesure où la loi le permet...dire que la volonté est autonome, c'est ignorer la hiérarchie des normes juridiques" ; Más favorable al principio, MALAURIE, Ph. Et AYNES, L., *Obligations. 2. Contrats et quasi-contrats*, Paris, Cujas, 2001/2002, nº 352, pp. 199-200: "Plus modérément, on doit mantenir le principe, tout en soulignant que la puissance de la volonté n'est pas absolue: elle se heurte aux réalités qui lui sont extérieures, et aux nécessités de l'organisation sociale";

Las precedentes explicaciones muestran el equívoco de considerar la autonomía de la voluntad un principio vigente en la obra de Domat, la codificación francesa y la temprana *école de l'exégèse*. Solo a finales del siglo XIX con la vulgarización de la obra de Kant, la doctrina jurídica francesa comienza a sentar las bases de este dogma para explicar la teoría general del contrato. El análisis de las fuentes del *Code* así como la evolución de la doctrina civil permite constatar que el fundamento de la fuerza obligatoria del contrato no estaba en la autonomía de la voluntad, concepto en ese entonces ignorado. Ahora, creo necesario intentar trazar la acogida de esta teoría en el derecho chileno.

7. **La recepción del dogma en la doctrina chilena.** El artículo 1545 del *Código civil* es, en parte, la traducción del artículo 1134 del *Code* Napoléon. No cabe duda que este último es la fuente inspiradora del primero, a pesar que en los diversos proyectos de *Código civil* no existe una cita explícita del precepto francés y ningún comentario de Bello al respecto.[43] Así lo reconoce la civilística nacional.[44] Las escasas obras jurídicas en el siglo XIX no aluden a la autonomía

[43] En el proyecto de Código civil (1841-1845) De los contratos i obligaciones convencionales, cuya publicación comienza en El Araucano nº 627 hasta el nº 800, el artículo 1º del título XI señala "Todo contrato legalmente formado es una lei para los contratantes, i no puede ser revocado sino por su consentimiento mutuo o por causas legales". (sin nota); En el Proyecto de Código civil (1846-1847), libro de los contratos i obligaciones convencionales, impreso en agosto de 1847, el artículo 92 del título XII señala "Todo contrato legalmente celebrado es una lei para los contratantes, i no puede ser revocado sino por su consentimiento mutuo o por causas legales" (sin nota). Salvo el cambio en la expresión "formado" por "celebrado" no existe ningún cambio en la redacción. Véase BELLO, A., *Obras completas. Proyecto de Código civil*, t. III, 1er tomo, Santiago, Nascimiento, 1932, p. 177 y 430; Por su parte el Proyecto de 1853, libro IV De los contratos i obligaciones convencionales, en su artículo 1724 del título XII consagra la redacción definitiva del precepto: "Todo contrato legalmente celebrado es una lei para los contratantes, i no puede ser invalidado sino por su consentimiento mutuo o por causas legales" (sin nota), *Idem*, t. IV, 2º tomo, p. 407; redacción que será ide´ntica en el denominado Proyecto inédito en su artículo 1724, *Idem*, t. V, 3er tomo, p. 411.
[44] Véase GUZMÁN BRITO, A., « Le Code de Napoléon et le code civil du Chili », en *La circulation du modèle juridique français*, Travaux de l'association Henri CAPITANT, t. XLIV, Paris, 1993, p. 151, quien afirma, de manera equivocada, que la libertad contractual "es coronada con el principio de la autonomía de la voluntad consagrado en el código chileno por su artículo 1545 correspondiente al artículo 1134 del francés"; CLARO SOLAR, L., *Explicaciones de derecho civil chileno y comparado. De las obligaciones*, Santiago, Nascimiento, 1937, t. 11, nº 749, p. 114 y nº 1029, p. 468; LÓPEZ SANTA-MARÍA, J., *op. cit.*, (n. 2), t. I, nº 45, p. 265; LEÓN HURTADO, A., *La voluntad y la capacidad en los actos jurídicos*, Santiago, Jurídica, 1979[3], p. 57, nota (1).

de la voluntad.[45] Esta expresión solo comienza a utilizarse en Chile en la obra de Claro Solar, aunque éste no incurre en la confusión de imputar a la autonomía de la voluntad el fundamento obligatorio del contrato, sino que la trata como libertad contractual.[46] Luego, de la misma manera que los autores franceses, la doctrina nacional que acoge la expresión autonomía de la voluntad y que la eleva a la categoría de principio, una vez expuesta de manera más o menos sistemática, da cuenta de los peligros de aceptar este dogma a ultranza. Es decir, al estudiar la autonomía de la voluntad describen su declive y realizan un análisis crítico de la misma.[47] Más que preocuparse de la autonomía de la voluntad como fundamento de la obligatoriedad del contrato, se alude a la necesidad de establecer límites a la misma ante los posibles abusos en su ejercicio. Sin duda, la doctrina nacional reconoce, siguiendo la doctrina francesa, el asiento del dogma de la autonomía de la voluntad en el artículo 1545 del *Código civil*. Esta conclusión es confirmada en la obra de Claro Solar, quien al referirse a la "autonomía de la voluntad" reproduce las explicaciones propuestas por la dogmática gala. En efecto, Claro Solar señala que "esta noción de la libertad individual tiene, en el dominio del derecho, un carácter más restringido y preciso bajo el nombre de principio de la *autonomía de la voluntad*".[48] En apoyo de esta idea cita a la obra de Planiol, Ripert y Esmein, Demogue, Colin y Capitant, Saleilles, Gény y Bonnecase. Estos autores franceses a que alude Claro Solar otorgan un valor dispar a la autonomía de la voluntad, sin reconocerle el valor de fundamento de la fuerza obligatoria del contrato.[49] En el mismo sentido, Claro Solar no justifica la fuerza obligatoria del contrato en la autonomía de la voluntad, sino que se refiere a ella en su dimensión de fuente generadora de obligaciones y, siguiendo a Demogue, analiza la oposición entre la teoría de la voluntad y la teoría de la declaración de voluntad.[50] Además, cuando trata la fuerza obligatoria del contrato, "este principio fundamental que da al contrato su eficacia", se preocupa de fundamentar la procedencia del recurso de casación por infracción del contrato, sin otorgar a la autonomía de la voluntad el valor de fundamento.[51] Solo en la década del 30, los autores nacionales comienzan a referirse a la autonomía de la voluntad como fundamento de la fuerza obligatoria del contrato. Somarriva, nos dice que el principio de la autonomía de la voluntad esta recogido en diversos preceptos del Código civil, entre los cuales, "en forma muy principal, el artículo 1545 del *Código civil*, según el cual el contrato válidamente celebrado es ley para los contratantes"[52] Y, luego, entre las consecuencias que se derivan del principio señala la fuerza obligatoria del contrato.[53] Por su parte, Alessandri, quien reconoce el principio de la autonomía de la voluntad,[54] al referirse al

Este último se remite a Claro Solar, de quien toma una cita de Saleilles.

[45] La doctrina chilena del siglo XIX no se refiere al principio de la autonomía de la voluntad. FÁBRES, J. C., *Instituciones de derecho civil chileno*, Valparaíso, imprenta del universo, 1863, nº 426, p. 182. Este autor se limita, siguiendo la temprana *Ecole de l'exégèse*, a comentar brevemente artículo por artículo. En la publicación de las *Institucione*s en sus *Obras completas*, se agrega una nota (62), en la cual solo se alude a la confusión entre efectos de las obligaciones y efectos del contrato, FABRES, J. C., *Obras completas. Instutciones de derecho civil chileno*, Santiago, La ilustración, 1912, p. 336. Por su parte, VERA, R., *Código civil de la República de Chile*, t. V, Santiago, Imprenta de la Gaceta, 1897, no se refiere ni utiliza la expresión autonomía de la voluntad.

[46] Agradezco los comentarios al profesor Alejandro Guzmán Brito en esta parte, a propósito de quien hago justicia a Claro Solar en no haber confundido el problema aquí tratado.

[47] CLARO SOLAR, L., (n. 44), nº 749, p. 115; SOMARRIVA UNDURRAGA, M., "Algunas consideraciones sobre el principio de la autonomía de la voluntad", en *RDJ*, t. XXXI, 1934, p. 37; LLANOS MEDINA, A., *El principio de la autonomía de las voluntad y sus limitaciones*, Memoria de Prueba, Universidad de Chile, 1944; HOJMAN PEZOA, B., *Autonomía de la voluntad y libertad contractual (ensayo crítico)*, Memoria de Prueba, Santiago, El chileno, 1945; PÉREZ GUERRERO, D., *El principio de la autonomía de la voluntad, artículo 1545 del Código civil*, Memoria de Prueba, Universidad de Chile, Santiago, Universitaria, 1958.

[48] CLARO SOLAR, L., (n. 44), nº 749, p. 113. Cfr. BARROS ERRÁZURIZ, A., *Curso de derecho civil*, Santiago, Imprenta Cervantes, 1921[3], p. 339, quien al referirse a los efectos jurídicos de los contratos no menciona la "autonomía de la voluntad" y se limita a comparar el contrato y la ley. Tampoco la menciona, DE LA MAZA R, L., *Derecho civil. Contratos*, Santiago, Universitaria, 1954, p. 138.

[49] GÉNY, F., (n. 39), nº 170 a 173. Este autor no la analiza como un principio sino como un problema jurídico y no le reconce el carácter de fundamento de la fuerza obligatoria del contrato; DEMOGUE, R., *Traité des obligations en général. Sources des obligations*, Paris, Russeau, t. I, nº 27 y ss. Demogue no se ocupa de la autonomía de la voluntad con relación a la fuerza obligatoria, sino que analiza el debate, siguiendo de cerca la tesis de Gounot, entre la teoría de la voluntad y la teoría de la declaración de voluntad, en particular vid. nº 32 y ss.; SALEILLES, R., *De la déclaration de volonté. Contribution à l'étude de l'acte juridique dans le code civil allemand*, Paris, LGDJ, 1929. Saleilles alude a la autonomía de la voluntad como regla dominante en materia de actos privados, lo que permite a las partes de crear todo tipo de convenciones que no sean contrarias a las buenas costumbres, al orden público o a las leyes, vid. nº 6, p. 196. Sin embargo, Saleilles es partidario de una regulación del ejercicio de la autonomía de la voluntad. Recuérdese que el mismo fue el creador de la expresión contrat d'adhésion con el objeto de limitar la libertad contractual, vid. nº 44, p. 214 y nº 10, p. 255.

[50] CLARO SOLAR, L., (n. 44), p. 118 y ss.

[51] *Op. cit.*, nº 1026, p. 465. Cabe señalar que Claro Solar no tuvo a la vista la tesis de Gounot, que, según dijimos, posibilitó la relación entre el dogma de la autonomía de la voluntad y la fuerza obligatoria del contrato. Esta puede ser la razón que no incurra en la impropiedad de asimilar la fuerza obligatoria a la autonomía de la voluntad. Vid. Corte Suprema, 12 de noviembre de 1926, en *RDJ*, t. XXIV, 1927, p. 289, nota CLARO SOLAR, en particular nº 15 y ss.

[52] SOMARRIVA UNDURRAGA, M., (n. 47), p. 39.

[53] *Ibidem*

[54] ALESSANDRI RODRÍGUEZ, A., *De los contratos*, Santiago, jurídica, s/f, nº 12, p. 10; "El contrato dirigido",

fundamento de la obligatoriedad de los contratos, señala que se trata de un problema que pertenece a la filosofía del derecho y, agrega, que existen importantes discrepancias al respecto. Transcribe las ideas de Ruggiero, según el cual, el fundamento de la obligatoriedad del contrato debe encontrarse en la *unidad de la voluntad contractual*.[55] La confusión que explica el fundamento de la fuerza obligatoria del contrato por la autonomía de la voluntad se encuentra en su obra. Este autor al referirse al efecto que produce el contrato entre las partes, alude al artículo 1545 del *Código civil* como "una disposición interesante del *Código civil* y que sirve mucho en la aplicación práctica. Contempla en forma clara y precisa el principio de la autonomía de la voluntad".[56] Con posterioridad, el pensamiento de Alessandri tendrá una importante influencia en la dogmática civil chilena. El reconocimiento de la autonomía de la voluntad en el artículo 1545 del *Código civil* pasará a ser usual en la literatura jurídica nacional.[57] Esta interpretación no resulta sorprendente, pues la civilística nacional ha sido influenciada de manera importante por la doctrina francesa.[58] Como puede observarse solo en el siglo XX la doctrina nacional comienza a utilizar la expresión autonomía de la voluntad y la amalgama entre esta y la fuerza obligatoria del contrato aparece por primera vez en la obra de Alessandri, confusión que tiene su origen en una lectura descuidada y descontextualizada de la doctrina francesa de los primeros decenios del siglo XX. Esta confusión esta presente en la civilística contemporánea. Así, Abeliuk afirma de manera categórica que "los códigos Civil y de Comercio, dictados en pleno auge internacional del principio que comentamos, lo recogen integramente...La norma fundamental es el ya citado Art. 1545 que otorga fuerza de ley a los contratos". Para Abeliuk, resulta evidente que la fuerza obligatoria del contrato deriva del principio de la autonomía de la voluntad.[59] López-Santa María sin plantear una posición sobre el problema, expone las diferentes doctrinas al respecto. En particular, resume la posición de Gounot, Giorgi y Ghestin y, además entrega una precisión metodológica. Este autor se distancia de la doctrina usual que asume la fuerza obligatoria como consecuencia ineluctable de la autonomía de la voluntad.[60] Por último, Guzmán Brito, explica que la voluntad no puede ser la causa de un derecho sino que solo puede dar lugar a otros hechos en el entendido que la voluntad constituye un fenómeno psicológico fáctico. Esta afirmación la realiza para descartar una "concepción subjetivista y voluntarista de los fenómenos jurídicos", la cual estima inadmisible.[61] Como puede observarse la doctrina nacional a partir de la obra de Alessandri afirma el dogma de la autonomía de la voluntad sin tener en cuenta los estudios dogmáticos e históricos realizados. No parece correcto continuar señalando que el artículo 1545 del *Código civil* acoge el principio de la autonomía de la voluntad. Dicho precepto según la tradición histórica alude a una cuestión distinta y que es la fuerza obligatoria del contrato. Esta a su turno no se justifica en la voluntad de las partes. Sin duda sería una paradoja que el homenaje del *Código civil* a la autonomía de la voluntad consista en imponer al deudor la perseverancia de su voluntad fundacional aunque esta haya manifiestamente cambiado. Cuando el artículo 1545 reproduce el artículo 1134 del Code afirmando que *« todo contrato legalmente formado es una ley para las partes »*, no se pronuncia sobre

en *RDJ*, t. XXXVIII, 1941, p. 5; ALESSANDRI RODRÍGUEZ, A. y SOMARRIVA UNDURRAGA, M., VODANOVIC H., A., *Tratado de las obligaciones. De las obligaciones en general y sus diversas clases*, Santiago, jurídica, 2001[2], nº 61, p. 47.

[55] ALESSANDRI RODRÍGUEZ, A. y SOMARRIVA UNDURRAGA, M., VODANOVIC H., A., *Curso de derecho civil. Fuentes de las obligaciones*, Santiago, Nascimiento, 1942, t. IV, nº 19, p. 21.

[56] ALESSANDRI RODRÍGUEZ, A. y SOMARRIVA UNDURRAGA, M., VODANOVIC H., A., (n. 54), nº 332, p. 208.

[57] Podemos citar las siguientes memorias de prueba de la universidad de Chile, LLANOS MEDINA, A., (n. 47), nº 31, p. 63 y 65: "El legislador, tomando en consideración la importancia que tiene este aspecto de la autonomía de la voluntad (la ley del contrato), en la vida práctica, ha tratado de establecer sobre una base sólida toda creación jurídica de la voluntad y a ese fin tiende aquella fórmula, que invariablemente encontramos en casi todas las legislaciones del mundo, según la cual todo contrato legalmente celebrado es una ley para los contratantes". Y luego agrega: "De cesta manera, el principio de la autonomía de la voluntad encuentra su más sólido afianzamiento en su aspecto positivo, en el reconocimiento que el legislador hace del carácter obligatorio de los contratos". Más adelante, siguiendo la línea trazada por Claro Solar y Alessandri, realiza una descripción de las transformaciones y atenuación del principio de la autonomía de la voluntad. *Idem*, p. 123 y ss. Lo copia sin citarlo, PÉREZ GUERRERO, D., *op. cit.*, (n. 46), p. 47. Cfr. HOJMAN PEZOA, B., (n. 47), nº 123, 124, p. 71 y ss. y p. 99: "La voluntad actúa como causa eficiente. Nadie lo niega pero, lo es en virtud de una función delegada, de una función que le otorga el derecho.".

[58] DÍAZ MUÑOZ, E., *El efecto relativo de los contratos*, Santiago, Jurídica, 1985, p. 14, el cual cita BAUDRY-LACANTINÉRIE et BARDE.

[59] ABELIUK MANASEVICH, R., *Las obligaciones*, t. I, Santiago, Jurídica, 2001[4], nº 102, p. 117: "Moralmente, el principio que comentamos (la fuerza obligatoria del contrato), heredero de la autonomía de la voluntad, encuentra su justificación en el aforismo "pacta sum servanda"".

[60] LÓPEZ SANTA-MARÍA, J., (n. 2), nº 13 y 45, p. 66 y 265.

[61] GUZMÁN BRITO, A., "Contribución a la crítica del dogma de la voluntad como fuente de efectos jurídicos", en Barros Bourie, E. (coordinador), *Contratos*, Santiago, jurídica, p. 209 y en particular pp.258-261.

el fundamento de la fuerza obligatoria del contrato. Es la doctrina que ha querido descubrir en esta frase la cosacración del principio de la autonomía de la voluntad, la cual explicaría, a su turno, la fuerza obligatoria de las convenciones.[62] Por cierto, negar el rol de fundamento a la autonomía de la voluntad de la fuerza obligatoria, no significa desconocer la relevancia de la "voluntad" como causa eficiente de los contratos, aunque no sea siempre así. Es indudable que la voluntad de las partes, sin considerarla como dogma o principio, es relevante para la formación del contrato o para la interpretación del contenido contractual. El contrato se forma, por regla general, por el encuentro de las voluntades de las partes, condición esencial para la existencia del contrato. Esta es la razón por la cual el artículo 1445 n° 2 del *Código civil* considera la voluntad exenta de vicios un requisito de validez del contrato. Pero una vez formado el consentimiento, la voluntad es impotente para explicar la fuerza obligatoria del contrato. Porque la voluntad de cada contratante pierde, una vez perfeccionado el contrato, su fuerza creadora o modificadora, ella no puede explicar la fuerza obligatoria del contrato. En otros términos, la voluntad de las partes constituye un elemento fundamental para la formación del contrato, mas carece de fuerza explicativa de su obligatoriedad. La voluntad de las partes determina, en principio, el contenido obligacional, esto es, las obligaciones que deberán cumplir los contratantes, pero no puede explicar el carácter obligatorio del contrato. Como lo muestra Kelsen, la autonomía de la voluntad no puede explicar la fuerza obligatoria del contrato, porque no existe razón que justifique preferir la voluntad al momento de la conclusión del contrato en desmedro de aquella vigente durante la ejecución del mismo.[63] No cabe confundir el contenido del contrato, cuya causa eficiente es la voluntad, y la explicación de la naturaleza coercitiva del vínculo contractual.[64] A decir verdad, el artículo 1545 del *Código civil* es una regla con contenido neutro, este precepto no dice que el contrato es obligatorio porque ha sido querido por las partes, solo afirma que el contrato legalmente celebrado es una ley para las partes. En este sentido el artículo 1545 es más explícito que su símil francés. El artículo 1545 no hace una elípsis para afirmar la naturaleza legal del vínculo obligatorio. El acuerdo de las partes da nacimiento a una norma jurídica la cual explica la fuerza obligatoria del contrato.[65] Por lo anterior, cabe concluir que el artículo 1545 del *Código civil* que constituye una ley en el sentido técnico del término, se limita a reconocer las fuerza obligatoria del contrato que ha sido el resultado del acuerdo de voluntades de las partes.

8. **Conclusión y perspectivas.** Las precedentes reflexiones muestran que la autonomía de la voluntad no permite explicar el fundamento obligatorio del contrato. Por el contrario, la concepción tradicional del artículo 1545 del *Código civil* confunde el rol que debemos atribuir legítimamente a la voluntad de las partes y el fundamento de la obligatoriedad del contrato. Si bien la voluntad de las partes puede explicar la formación del contrato, una vez celebrado, aquella resulta insuficiente para explicar el carácter normativo del mismo. En definitiva, la autonomía de la voluntad no constituye el fundamento de la fuerza obligatoria del contrato. Esta conclusión podría ser pasablemente teórica, aunque presenta de por sí el interés de re-construir la historia jurídica de este principio que ha modelado toda la teoría general del contrato en los tiempos modernos. Con todo, la concepción voluntarista del contrato implica consecuencias jurídicas importantes. Si bien no corresponde tratarlas aquí, cabe tener presente que el dogma de la autonomía de la voluntad justificaría no solo la fuerza obligatoria del contrato, sino también el efecto relativo de los mismos. La fijación de quiénes deben ejecutar el contrato se encuentra marcada por el dogma de la autonomía de la voluntad. Plantear una concepción distinta de la fuerza obligatoria del contrato, ya no basada en la voluntad de quienes concurren a celebrar el contrato, podría posibilitar la introducción de la figura de los grupos

[62] WEILL, A., *Le principe de la relativité des conventions en droit privé français*, thèse, Strasbourg, Dalloz, 1938, 1062 p. 3 y ss..
[63] KELSEN, H., « La théorie juridique de la convention », en *Arch. Ph. de droit*, 1940, p. 33.
[64] Sobre esta distinción, Vid. ANCEL, P., (n. 1), p. 771 y ss. Este autor muestra de manera precisa la distinción entre la fuerza obligatoria del contrato como resultado de una norma jurídica convencional y el contenido obligacional del contrato.
[65] En este sentido, GOUNOT, E., (n. 1), p. 342 y sis.; ROUHETTE, G., (n. 1), p. 398; RIEG, A., « Le contrat dans les doctrines allemandes du dix-neuvième siècle », en *Arch. Ph. du droit*, 1968, p. 31 ; BACACHE-GIBEILI, M., (n. 3), n°283, p. 248 ; STARCK, B., ROLAND, H. et BOYER, L., *Droit civil. Les obligations. 2. Contrat*, 1998[6]. n° 18, p. 7.

de contratos. En efecto, el dogma de la autonomía de la voluntad delimita al mismo tiempo las fronteras entre la responsabilidad contractual y la responsabilidad extracontractual. Sin embargo, la presencia de grupos de contratos como nueva categoría contractual plantea el desafío de realizar una re-lectura del efecto relativo de los contratos, ya no anclada en el dogma de la autonomía de la voluntad y posibilitar la extensión de la noción de parte contractual a fin de estructurar el régimen jurídico de la acción directa al interior de los grupos de contratos. Esta re-lectura del efecto relativo de los contratos, alentada por esta crítica al dogma de la autonomía de la voluntad, podrá justificar la introducción de la noción de grupos de contratos. La teoría general del contrato anclada en el dogma de la autonomía de la voluntad aparece desmentida por el análisis de las fuentes legales y por la evolución de la contratación. Esta crisis del voluntarismo requiere responder a la pregunta "La teoría general del contrato, Mito o realidad?.[66]

[66] SAVAUX, E., *La théorie générale du contrat, mythe ou réalité?*, Préface J.L. Aubert, Paris, LGDJ, 1997.

CONTRACT LAW, DEFAULT RULES, AND THE PHILOSOPHY OF PROMISING

*Richard Craswell**

Among the topics addressed by moral philosophy is the obligation to keep one's promises. To many philosophers, there is something strange (or, at least, something calling for explanation) in the idea that moral obligations can be created simply by an individual's saying so — yet this is what seems to happen when a person makes a promise. Consequently, there is by now a large body of literature attempting to identify the exact source and nature of this moral obligation.

Contract law, too, has something to do with promises, so philosophers of law (and philosophically minded lawyers) often draw on philosophical theories about promising when writing about contract law. For example, a recent book by Charles Fried purports "to show how a complex legal institution, contract, can be traced to and is determined by a small number of basic moral principles . . . ."[1] In Fried's view, a recognition of the proper philosophical basis of contract law leads to conclusions profoundly different from those that would result from any attempt to rest contract law on other social policies, such as economic efficiency or the redistribution of wealth.[2]

My thesis is that such claims on behalf of philosophical theories of promising are greatly exaggerated. In particular, analyses such as Fried's have little or no relevance to those parts of contract law that govern the proper remedies for breach, the conditions under which the promisor is excused from her duty to perform, or the additional obligations (such as implied warranties) imputed to the promisor as an implicit part of her promise. These doctrines, which serve to define the exact scope of contractual obligations, are often referred to as

* Professor of Law, University of Southern California. B.A. 1974, Michigan State University; J.D. 1977, University of Chicago. — Ed. I have benefited from comments by Scott Altman, Ian Ayres, Randy E. Barnett, Richard A. Epstein, John Finnis, John Gardner, Ronald R. Garet, Catharine W. Hantzis, Michael S. Moore, Joseph Raz, Alan Schwartz, W. David Slawson, Christopher D. Stone, and participants in workshops at USC and at Oxford University. I am also grateful for financial support from the USC Law Center Summer Research Fund and the USC-Oxford Legal Theory Institute.

1. C. FRIED, CONTRACT AS PROMISE: A THEORY OF CONTRACTUAL OBLIGATION (1981), at the first (unnumbered) page of the preface. *See also* Barnett, *Contract Scholarship and the Reemergence of Legal Philosophy* (Book Review), 97 HARV. L. REV. 1223 (1984) (applauding what he sees as the growing recognition of the importance of normative philosophy to contract law and to legal scholarship generally).

2. C. FRIED, *supra* note 1, at 4-6, 74-85, 103-11.

"background rules" or "default rules," although the term "default rules" more commonly refers only to those rules which the parties are free to vary by appropriate language in their contract.[3] As not all contract rules can be varied in this way, I will use the term "background rules" to refer to both waivable and nonwaivable rules.

I am not asserting here that philosophical theories about promising can have no implications for any part of contract law. Such theories may well have implications for questions about the proper scope of freedom of contract — that is, questions about whether any given rule ought to be merely a default rule, or whether it ought to be mandatory for all parties. In addition, *certain* philosophical theories may have implications for the proper content of contract law's background rules. For example, theories that justify the enforceability of promises on grounds of economic efficiency, or on the special value of certain kinds of relationships, may imply that the law should adopt those background rules that are most efficient or that best promote the most highly valued kinds of relationships.

Other philosophical theories, however — including the one endorsed by Fried — have no such implications for the content of the law's background rules. These theories ground the enforceability of promises on considerations of individual freedom and autonomy, or on the principle of fidelity to one's prior statements or commitments. In a nutshell, the fidelity principle is consistent with any set of background rules because those rules merely fill out the details of what it is a person has to remain faithful to, or what a person's prior commitment is deemed to be. Thus, while fidelity may dictate that a promisor must live up to the obligations described by any set of background rules the law has adopted, it cannot guide the legal system in deciding which background rules to adopt in the first place. The principle of individual freedom is equally unhelpful, for it implies only that individuals should be left free to change whatever default rule the law adopts as a starting point. Once again, some other value must be invoked to explain why one starting point ought to be picked by the law in preference to another.

If I am right, this means that it is not enough to reject notions such as economic efficiency, or theories that value certain kinds of relationships more highly than others, as insufficient or incorrect justifications for the basic proposition that promises ought to be enforced. Even if

3. The role of "default rules" is discussed at more length in Goetz & Scott, *The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms*, 73 CALIF. L. REV. 261 (1985), and Ayres & Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 YALE L.J. 87 (1989).

those theories are rejected at this most basic level, further argument is needed to explain why they are inadmissible bases for selecting the law's background rules or starting points, and, if so, to explain what other values ought to be used instead. Unfortunately, current writings bringing philosophical analysis to bear on contract law have focused almost entirely on the more basic question of why promises ought to be enforced at all, and have tended to overlook this second level of analysis. This frequently leads to careless or ad hoc statements concerning the proper content of contract law's background rules.

Part I of this article presents a more detailed survey of recent philosophical writings about promises, for the benefit of legal readers who may be unfamiliar with that literature. Part II then discusses the role of background rules in contract law, and shows why the content of those rules cannot be derived from philosophical theories based on individual liberty, or on ideals such as fidelity or truthfulness. Finally, Part III examines the writings of Charles Fried and Randy Barnett to illustrate the consequences of attempting to apply philosophy to contract law without addressing these problems. These two authors have supplied the most comprehensive attempts to give contract law a philosophical grounding, yet each falls into exactly this error.

I. Philosophical Theories About Promising

Broadly speaking, there are two kinds of questions that can be asked about the ethics of promising. One category consists of questions that might be asked by someone wondering whether to make a promise — for example, "Should I promise to donate money to the poor?" or "Should I promise to help a friend, if I can do so at little inconvenience to myself?" The other consists of questions facing someone who has already made a promise — for example, "Should I carry out my promise, even though my circumstances or my desires have changed?" The first class of questions asks what kinds of promises ought to be made, while the second asks what follows from having made a promise.

The first class of questions has received little attention in philosophical writings about promising as such. This should not be surprising, for the ethical theories most relevant to questions about what kind of promises to make will usually be theories with implications far beyond the topic of promising. If a person is wondering whether to promise money to the poor, the most interesting question (from the standpoint of ethics) is whether she ought to help the poor at all. The subsidiary question of whether she ought to help them by *promising* money, rather than giving them money without ever having first

promised to do so, seems much less important. Thus, ethical theories about what kind of promises to make usually derive from theories about the particular subject matter of the promise (helping the poor, etc.). They do not derive from theories about promising as such.

The same could be true of the second class of questions, concerning the ethical consequences of having made a promise. That is, there would be nothing illogical in believing that the conditions under which it is excusable to break a promise to the poor have no connection (in the sense of being linked by any common theory) with the conditions under which it is excusable to break a business promise, or a promise to a friend. If that were the case, there would be no point in asking questions about the nature of the commitment represented by promises in general. One could speak of the commitment represented by charitable promises, or business promises, but it would be useless to search for any general, unifying theory of promises.

Most people, though, would reject the idea that the obligations imposed by different kinds of promises have nothing in common. The mere fact that we classify certain speech acts as "promises" strongly suggests that they have something in common; otherwise, there would be no point to that classification. Indeed, there is a growing body of philosophical literature which attempts to describe the obligation someone accepts when she promises to do some action ø, regardless of how ø is filled in. That literature presupposes that there are at least some things that can be said about promises without discussing the specific subject matter of the promise.

Contract law, too, is traditionally concerned with the elements that all promises have in common, so it is this literature that is most often invoked by legal scholars in search of a philosophical grounding for contract law. The remainder of this Part describes this literature in more detail. One branch of the literature takes it as given that promises are morally binding and attempts to describe the exact way in which this "bindingness" should constrain the promisor's subsequent conduct. Another branch attempts to articulate the reasons why promises might be morally binding. While these two inquiries are related at many points, it will be convenient to discuss them separately.

A. *What Does It Mean To Be Bound by a Promise?*

Assume that a person has promised to do some action ø. Most people would agree that this does not place the promisor under an absolute duty to do ø — for example, if ø becomes impossible for reasons beyond the promisor's control, she may be excused from her obli-

gation.[4] Most people would also agree that the promisor's freedom has been constrained in some way, so she cannot decide whether or not to do ø with the same freedom she would have had if she had not made the promise. If asked to articulate the exact way in which her freedom is constrained, however, most people would have a great deal of difficulty.

Philosophers have taken two somewhat different approaches to articulating the way in which a promise constrains the promisor. One approach views the promise as creating additional reasons in favor of doing ø; the other views the promise as barring the promisor from considering certain reasons that might otherwise argue against doing ø. Each of these will be discussed below.

1. *Promises as Excluding Reasons for Action*

One view of the way that promises constrain a promisor's subsequent deliberations can be found in a 1955 article by John Rawls.[5] Rawls was only indirectly concerned with promises, as his main objective was to describe the distinction between justifying a social practice and justifying an individual action within such a practice. But because Rawls found the practice of promising to be a useful example, much of what he said is relevant here.

Rawls' point was that some practices, including promising, are defined so as to make certain arguments no longer available to those acting within the practice. More specifically, once a person has promised to do something, it is no longer open to that person to decide not to perform on the ground that, all things considered, nonperformance seems preferable to performance.[6] While such a preference might be a perfectly proper ground on which to refuse to make a promise in the first place, the rules of promising foreclose such an argument once the promise has been made.

Rawls recognized that this did not mean that a promisor was obliged to carry out her promise regardless of the circumstances:

> Is this to say that in particular cases one cannot deliberate whether or not to keep one's promise? Of course not. But to do so is to deliberate whether the various excuses, exceptions and defenses, which are understood by, and which constitute an important part of, the practice, apply to one's own case. Various defenses for not keeping one's promise are allowed, but among them there isn't the one that, on general utilitarian grounds, the promisor (truly) thought his action best on the whole, even

4. For convenience in the use of pronouns, all my examples will assume a female promisor and a male promisee.

5. Rawls, *Two Concepts of Rules*, 64 PHIL. REV. 3 (1955).

6. *Id.* at 16-17.

> though there may be the defense that the consequences of keeping one's promise would have been *extremely* severe.[7]

Rawls argued that the practice of promising must rule out an excuse based on a general balance of considerations — otherwise, the point of having the practice would be lost.

More recently, Joseph Raz has developed a similar theory in which promises supply what he calls "exclusionary reasons."[8] According to Raz, a promise to do ø constrains the promisor's options by preventing her from giving any weight to certain arguments that might be relevant if she had not promised. In other words, promises are not themselves reasons for doing ø, but they bar or exclude certain factors which might otherwise be reasons *not* to do ø.

Like Rawls, Raz recognizes that promises do not exclude *all* other considerations, thereby leaving the promisor under an absolute obligation to do ø. If ø would involve actions that were themselves immoral, for example, it might still be permissible for the promisor to choose not to do ø.[9] Indeed, Raz does not attempt to list the precise reasons that are excluded by a promise — though he would at least agree with Rawls that the reason, "all things considered, it seems best not to ø" is one of the excluded reasons.[10] A promise binds a promisor, in Raz's view, by excluding such arguments from the promisor's subsequent calculations.

2. *Promises as New Reasons for Action*

A different class of theories holds that a promise creates new reasons to perform the promised action, which must then be added to any pre-existing balance of reasons for and against the action. For example, a promise may give rise to expectations in the promisee, and the fact that nonperformance would disappoint those expectations may count as a reason favoring performance.[11] If the promisee has relied on the promise in any way, so that nonperformance would leave him worse off than if the promise had never been made, this might provide

7. *Id.* (footnote omitted).

8. *See especially* Raz, *Promises and Obligations*, in LAW, MORALITY AND SOCIETY 210, 222 (P. Hacker & J. Raz eds. 1977). For more general discussions of the concept of "exclusionary reasons" (or "pre-emptive reasons," as he refers to them in his later work), see J. RAZ, THE MORALITY OF FREEDOM 42-69 (1986); J. RAZ, PRACTICAL REASON AND NORMS ch. 1 (1975); Raz, *Authority and Consent*, 67 VA. L. REV. 103 (1981).

9. Raz, *Promises and Obligations*, *supra* note 8, at 211.

10. *Id.* at 222-23.

11. For examples of this theory, see Ardal, *And That's a Promise*, 18 PHIL. Q. 225, 233-37 (1968); Narveson, *Promising, Expecting, and Utility*, 1 CANADIAN J. PHIL. 207, 213-20 (1971). *See also* Downie, *Three Accounts of Promising*, 35 PHIL. Q. 259, 263-64 (1985) (attributing this argument to Adam Smith).

an even stronger reason in favor of performing the promise.[12]

As in the case of the theories discussed in the preceding subsection, these theories do not imply that a promisor must *always* carry out the promise. While the fact that the promisee would be disappointed may be a reason in favor of doing ø, nothing in this theory rules out the possibility of there being other, stronger reasons against ø. If ø is an immoral action, or has become impossible, the reasons against doing ø might well outweigh the reasons in favor of doing ø. Thus, both sets of theories are consistent with the belief that a promise is not binding in all circumstances.

Similarly, both sets of theories are consistent with the belief that a promise somehow militates in favor of performance, in the sense of requiring the promisor to choose ø in at least some cases where, absent the promise, she would otherwise have chosen not to do so. The first set of theories explains such cases by saying that the promise excludes certain reasons that might otherwise have counseled against doing ø, so the promise will have a decisive effect in any case where those excluded reasons would otherwise have been dispositive. The second set explains the promisor's decision by saying that all the old reasons for and against ø remain relevant, but the promise creates new reasons in favor of ø. On this view, the promise will make a difference whenever the balance of old reasons would have counseled against ø, but the addition of the new reasons is enough to tip that balance in favor of ø.

Finally, neither of these theories attempts to explain *why* a promisor is morally bound to limit her subsequent behavior in the way postulated by the theory. Instead, the goal of these theories is simply to illuminate the ordinary understanding of what it means to be under a promissory obligation, by explaining precisely what it is that a promisor is thought to be bound to do. The question of whether (or why) anyone is morally obliged to follow that ordinary understanding is a separate question which requires a separate analysis. Accordingly, the following section surveys some recent philosophical theories addressed to the question of why the rules of promising are or ought to be morally binding.

B. *Why Should Anyone Obey the Rules of Promising?*

The question of why the rules of promising are morally binding is easily confused with the question addressed in the preceding section,

12. Theorists who focus on the promisee's reliance, rather than the bare expectation of performance, include MacCormick, *Voluntary Obligations and Normative Powers*, 46 ARISTOTELIAN SOCY. 59, 62-63 (Supp. Vol. 1972), and Hanfling, *Promises, Games and Institutions*, 75 PROC. ARISTOTELIAN SOCY. 13, 15-18 (1975).

concerning exactly what it is that the rules of promising require. Some of this confusion was unintentionally introduced by a 1964 article by John Searle, provocatively entitled *How to Derive "Ought" from "Is."*[13] Searle's point was that certain kinds of descriptive assertions, or "is" statements, are statements describing "institutional facts." For example, the statement that someone has hit a home run — as distinct from the statement that someone has hit a round object over a fence with a piece of wood — makes sense only within the institutional framework created by the rules of baseball.[14] Similarly (in Searle's view), the statement that someone has made a promise makes sense only within the framework created by the rules of promising. Thus, to say that a person has promised is to describe that person with reference to the rules of promising, and those rules include the normative or "ought" statement that people who promise are bound to act differently in some way as a result of their promise.[15] In this sense, Searle argued, it is possible to move from certain kinds of "is" statements to certain kinds of "oughts."

A number of writers responded by pointing out that this argument, standing alone, does not really show that anyone who promises ought therefore to accept the rules of the practice of promising, or ought to regard those rules as morally binding.[16] As Searle later clarified, the only kind of "ought" that he was discussing was one that was internal to the practice of promising.[17] It would still be open for someone to reject the entire practice of promising, or to argue that some or all of that practice's rules were unjust. Searle's derivation of "ought" from "is" was never intended to supply the answer to this kind of "ought" question.

In consequence, writers have had to look elsewhere for the source of the moral obligation to respect the rules of promising. Here, too, the various explanations can be grouped into two categories, which correspond in some ways to the two theories discussed in the preceding section concerning what the rules of promising actually require. Some writers have argued that the obligations created by a promise can only be explained by positing that individual promisors possess

13. 73 PHIL. REV. 43 (1964).

14. *Id.* at 54-55.

15. *Id.* at 55-56.

16. *See, e.g.*, Carey, *How to Confuse Commitment with Obligation*, 72 J. PHIL. 276 (1975); Hare, *The Promising Game*, 70 REVUE INTERNATIONALE DE PHILOSOPHIE 398 (1964); Jones, *Making and Keeping Promises*, 76 ETHICS 287 (1966); Miller, *Constitutive Rules and Essential Rules*, 39 PHIL. STUD. 183 (1981); Robins, *The Primacy of Promising*, 85 MIND 321, 329-30 (1976).

17. J. SEARLE, SPEECH ACTS 188-89 (1969).

"norm-creating powers," under which they are authorized to create new moral obligations merely by agreeing to do so. Others have argued that no such powers are necessary, and that the moral obligation to keep a promise is merely a particular instance of a more general obligation, such as the obligation not to cause harm to others or the obligation to tell the truth. Each theory will be discussed below.

1. *Promises and Norm-Creating Powers*

Joseph Raz has offered the most complete defense of the position that the moral obligation to respect the rules of promising must rest on some norm-creating power possessed by the promisor.[18] Without being quite as explicit, and without necessarily agreeing with other aspects of Raz's theory, other writers have taken an essentially similar position. For example, Charles Fried has argued that notions of individual freedom and autonomy require that individuals be allowed to bind themselves by promising.[19] Economists have pointed to the social utility of allowing individuals to bind themselves to a future course of conduct, to make it easier for others to arrange their lives in reliance on the promise.[20] In Raz's terms, all of these are arguments that individuals ought to have at least one norm-creating power: the power to create a moral obligation by making a promise.

In a somewhat similar vein, Randy Barnett has argued that promises are best viewed as marking the promisor's consent to the transfer of part of her bundle of property rights.[21] "Property rights" are meant here in the broadest possible sense; they thus include future rights such as "the right to fifty bushels of wheat on August 1," or even disjunctive rights such as "the right to fifty bushels of wheat or their equivalent in money." On this view, if a seller has consented to transfer to a buyer the right to "fifty bushels of wheat on August 1," the seller no longer has any right to control those bushels when August 1 arrives, so her retention of those bushels would constitute the taking of another's property. Barnett's account of promising might seem more parsimonious than Raz's, as Barnett grounds the force of a promise in an already recognized power — the power to transfer property — rather than having to posit a separate normative power for promising. Notice, though, that the power to transfer title to property is itself an instance of what Raz would term a norm-creating power. In this sense, Barnett's theory is merely a particular example of the kind of theory that Raz believes necessary to account for the moral force of the rules of promising.[22]

18. Raz, *Voluntary Obligations and Normative Powers*, 46 ARISTOTELIAN SOCY. 79 (Supp. Vol. 1972). For a somewhat similar position, see Robins, *supra* note 16.

19. C. FRIED, *supra* note 1, at 14-17.

20. *See, e.g.*, Goetz & Scott, *Enforcing Promises: An Examination of the Basis of Contract*, 89 YALE L.J. 1261 (1980).

21. Barnett, *A Consent Theory of Contract*, 86 COLUM. L. REV. 269 (1986). As Barnett notes, this view of contract law dates back at least as far as Blackstone's Commentaries. *Id.* at 292 n.98; *see also* P. ATIYAH, PROMISES, MORALS, AND LAW 179 (1981) (attributing this view of promising to Hobbes).

Of course, positing a norm-creating power of this sort merely pushes the normative question back one step: What justifies the existence of this norm-creating power? To this question, many different answers can be and have been given. Economists usually rely on utilitarian justifications, arguing that a rule letting people make enforceable promises is likely (under certain conditions) to lead to the best use of resources. As noted above, Fried justifies the power to bind oneself as a corollary of respect for autonomy and individual freedom. Raz tentatively suggests a somewhat different answer: since promises create a special relationship between the promisor and promisee, the power to create such relationships is morally justified only if those relationships are themselves desirable.[23]

I will argue below that, when it comes to deriving legal implications, it may well make a difference which argument is used to justify these norm-creating powers.[24] In the philosophical literature on promising, though, these differences are less important (and receive less discussion) because all of these arguments lead to the same conclusion: that individuals ought to have the norm-creating power represented by promising. Thus, all of these theorists find themselves on one side of a divide, with the other side occupied by the theorists to be discussed in the next two sections.

2. *Promises and Harm to Others*

A different approach to explaining the moral force of the rules of promising begins with the observation that breaking a promise will

22. Raz occasionally distinguishes between consenting and promising, but his examples suggest that he is simply using "consent" in a sense that is narrower than Barnett's usage. For example, Raz argues that one can consent to give someone a right to some action, but cannot consent to give someone a mere right of recipience, such as the right to be paid a sum of money. Raz, *Authority and Consent*, *supra* note 8, at 121-22. Thus, where Barnett would say that a tenant has consented to transfer to her landlord the right to $1000 per month for each of the next 12 months, Raz would simply say that the tenant has promised to pay that sum. As nearly as I can tell, nothing turns on this difference in usage.

23. Raz, *Promises and Obligations*, *supra* note 8, at 227-28. In another context (while discussing consent generally), Raz has also noted the traditional instrumental arguments that letting people make their own promises is likely to lead to the best use of resources, or that promises develop peoples' characters and train them in such useful virtues as cooperation and careful planning. Raz, *Authority and Consent*, *supra* note 8, at 123-25.

24. *See infra* notes 48-57 and accompanying text.

often harm the promisee, especially if he has relied on the promise in any way. If there is a general principle that one ought not cause harm to others, that might be enough to justify some sort of rule against promise-breaking, at least when breaking the promise would in fact cause harm. In particular, this justification would fit well with the view that the only obligation imposed by a promise is an obligation to take this harm into account, as one reason — not necessarily the only one, or even a dispositive one — in favor of performing the promised action.[25]

Most people, though, are unwilling to equate a general obligation not to harm others with the specific obligation imposed by a promise. Imagine, for example, that a complete stranger walks up to you and says that he has formed the belief that you are about to give him $50,000 — and, moreover, that he has relied on this expectation by incurring various debts and obligations, none of which he would have incurred had he not believed that you were about to give him the money. The stranger's predicament in this case might provide a slight reason for you to help him out by giving him the $50,000 — after all, helping a fellow citizen is almost always a good reason for action — but it will not be a very strong one. Moreover, none of the writers cited above would say that preventing harm to the stranger is as strong a reason for action in this case as it would be if you had *promised* to give him $50,000, and your promise was what led him to incur his additional debts. Thus, some further argument is necessary to explain the particular force of the reason for action that is generated once a promise has been made.[26]

Some writers have tried to explain the additional force generated by a promise by pointing to the fact that a person who makes a promise has thereby *caused* the rise in the other party's expectations, and has likewise been a cause of the other party's reliance on the promise.[27] By contrast, the example in the preceding paragraph involved expectations and reliance which could not be causally attributed to any action of the person charged with the obligation. Thus, these writers rest the moral force of promissory obligations on a narrower form of the harm principle. While there is no obligation, or at best a very weak obligation, to take steps to *prevent* harm to others (these

25. *See supra* note 12.

26. For other criticisms of the bare-reliance argument, using examples similar to the one in the text, see C. FRIED, *supra* note 1, at 10-11; Raz, *Promises in Morality and Law* (Book Review), 95 HARV. L. REV. 916, 924-25 (1982).

27. *E.g.*, MacCormick, *supra* note 12, at 66-67.

writers would say), there is a much stronger obligation not actively to *cause* that harm.

Even this principle may be too broad, however. As a number of writers have pointed out, it is easy to imagine cases where one person has *caused* another to rely in some way, but the resulting obligation does not seem as strong as if the first person had made a promise to the second. Raz suggests the example of someone who informs a friend that she will almost certainly be able to offer the friend a ride into town, and that the likelihood of her being able to do so is sufficiently high that it makes perfect sense for the friend to rely on her rather than making alternative arrangements — but who warns, "remember — I do not promise anything, I am merely advising you."[28] In such a case, the friend presumably would be perfectly justified in relying on the statement by forgoing any alternative arrangements. The friend's lack of alternative arrangements may even be a factor that the speaker ought to consider in subsequently deciding whether to change her plans in some way that would prevent her from offering the ride. However, the force of this reason for going through with the offered ride still does not seem as strong as if the speaker had offered the ride and said, "I promise I'll be there." Thus, a promise still seems to add something to the force of the reasons for action over and above the force that can be attributed to the principle of not causing harm to others.

A closely related objection notes that the notion of "cause" employed in this argument is more than a little problematic.[29] Reliance on a promise is always caused (in a "but for" sense) by the promisee as well as by the promisor, for it is the promisee who chooses to rely. Thus, some further argument is needed to explain why we quickly attribute moral responsibility for another person's reliance to someone who has *promised* a particular action, but are less quick to attribute as much responsibility to someone who has made clear that she is not promising anything. To say that the person who relies assumes the risk of being disappointed if the other person's statement falls short of being a promise, but does not assume that risk if the other person has made an actual promise, is no answer. Such a response simply posits a difference in the extent of the speaker's responsibility in each case, when it is precisely that difference that needs a moral justification.[30]

28. Raz, *supra* note 18, at 99.

29. For an exceptionally clear statement of this objection, see P. ATIYAH, *supra* note 21, at 63-69; *see also* G. WARNOCK, THE OBJECT OF MORALITY 98-100 (1971); Robins, *supra* note 16.

30. Occasionally, the objection discussed in this paragraph is made by asserting that we cannot determine whether the promisee's reliance is reasonable without knowing whether the promise is binding on the promisor. *E.g.*, Barnett, *supra* note 21, at 275. However, Raz's example of

Patrick Atiyah suggested that the greater responsibility of a promisor might be explained by treating the promise as a conclusive admission that responsibility properly rested on the promisor. That is, Atiyah viewed the allocation of responsibility for the promisee's reliance as a question of social policy, which would normally have to be settled by a court or legislature or other lawmaking institution.[31] He then suggested that if an individual promisor *admits* that she should be responsible for the other party's reliance, by making an explicit promise, that sort of admission against interest should usually be treated as conclusively settling the social policy issue.[32] Notice, though, that this argument grants to individual promisors exactly the sort of norm-creating power for which Joseph Raz has argued.[33] If individual promisors have the sovereign-like authority to determine conclusively who should bear responsibility for certain losses, they are exercising a power whose existence must be justified by some principle beyond the general notion of not causing harm to others.

3. *Promises and Misrepresentation*

The attempt to explain why a speaker's responsibility seems greater if she has made a definite promise — even if the harm to the other party is no greater than in an otherwise similar case where no promise has been made — is one of the factors that led many writers to posit a separate, norm-creating power which individuals can exercise by making a promise. Other writers, however, have responded by narrowing the harm principle still further, saying that there is a particularly strong obligation not to cause harm to others *by making false statements.* Alternatively, the "harm" element can be eliminated entirely, if there is an obligation not to misrepresent the truth regardless of whether the misrepresentation causes any actual harm. Under either of these views, which I will refer to collectively as the "misrepresentation theory," the key fact is not that breaking a promise causes harm, but that breaking a promise violates the obligation to tell the truth.[34]

one friend informing another of a possible ride into town is enough to refute this proposition, by showing that there can be cases where it is reasonable for the promisee to rely regardless of whether the promise is binding. For further discussions of this issue, see McNeilly, *Promises Demoralized,* 81 PHIL. REV. 63 (1971); Narveson, *supra* note 11. Assertions such as Barnett's are probably better interpreted as making the argument discussed in the text: that we cannot know whether it is reasonable *to attribute responsibility for the reliance to the promisor* without first deciding whether the promise is binding.

31. P. ATIYAH, *supra* note 21, at 68-69.

32. *Id.* at 184-202.

33. Raz, *supra* note 26, at 926-27.

34. For examples of this approach, see G. WARNOCK, *supra* note 29, at 101-11; Ardal, *Ought*

The misrepresentation theory reaches this conclusion by interpreting all promises as representations about the promisor's future conduct. A person who promises to give a friend a ride has made a definite statement about what will happen in the future. If she then fails to come through with the ride, her failure makes this statement about the future a false one, thereby violating the obligation to tell the truth (in addition to causing harm). By contrast, a person who says that she will probably give her friend a ride, but who reserves the right to change her mind, has made a much weaker probabilistic statement about the future. That person's failure to perform would not make her previous statement false, and therefore would not violate the obligation to tell the truth, even though it might cause the same amount of harm to her friend. According to the misrepresentation theory, this is why a person who promises to give her friend a ride has a stronger reason to do so than does a person who has merely said that she is likely to give her friend a ride, without actually promising.

Admittedly, the obligation to tell the truth appears here in what may be an unfamiliar guise. The obligation has some elements of strict liability, for it is no defense to say that at the time of promising the speaker *thought* that her statement would come true.[35] In addition, the obligation to tell the truth is perhaps more usually thought of as an obligation limiting what one is allowed to say, by obliging people not to say anything that is false. Under this theory, though, the obligation limits what speakers can *do,* by forbidding people from doing anything that will make their prior statements turn out to be false. Rather than requiring people to conform their statements to reality, the misrepresentation theory of promising requires people to act in such a way that reality will conform to their prior statements.[36]

In some respects, then, the misrepresentation theory is not that different from theories based on the notion of norm-creating powers, as discussed above. If the obligation to tell the truth means that an individual, by choosing to making a definite statement about the future, can thereby place herself under an obligation to make sure that her statement comes true, then that individual certainly has the power to create an obligation of some sort. This obligation need not be of the sort discussed above in section I.A.1, however, which obliged the

We To Keep Contracts Because They Are Promises?, 17 VAL. U. L. REV. 655 (1983); Ardal, *supra* note 11, at 225; Fogelin, *Richard Price on Promising: A Limited Defense,* 21 J. HIST. PHIL. 289 (1983); Hanfling, *supra* note 12, at 24-25.

35. Compare the position of Charles Fried, discussed *infra* at note 68.

36. For a more extended discussion of this aspect of the theory, see Fogelin, *supra* note 34, at 296-97.

speaker not to consider certain reasons which might otherwise counsel against doing the promised act. The misrepresentation theory is equally consistent with obligations of the sort described in section I.A.2, which merely add the value of telling the truth as an additional reason in favor of performing the promised action.

II. The Needs of Contract Law

The rules of contract law can be divided into two categories: "background rules" and "agreement rules." As discussed earlier, background rules define the exact substance of a party's obligation, by specifying (among other things) the conditions under which her nonperformance will be excused, and the sanctions which will be applied to any unexcused nonperformance. By contrast, agreement rules specify the conditions and procedures the parties must satisfy in order to change an otherwise applicable background rule. Agreement rules thus include most of the rules governing offer and acceptance, as well as such doctrines as fraud or undue influence, which define the conditions necessary for a party's apparent consent to be counted as truly valid.[37] To be sure, these two categories are not mutually exclusive, for some rules serve both functions simultaneously. For example, the "mailbox rule" defines a procedure by which parties can agree to a contract changing any otherwise applicable background rule. It also provides a background rule of its own, by providing that the offeror's power to retract her offer ends as soon as an acceptance is posted (unless the offer itself provided otherwise). Notwithstanding such instances of overlap, the distinction between background rules and agreement rules is still useful in understanding the relationship between the philosophical theories of promising and the content of contract law.

The philosophical theories discussed above may well have some relevance for contract law's agreement rules. For example, theories that ground the enforceability of promises on individual liberty might argue that the parties should be allowed to overturn nearly all of contract law's background rules by an appropriate agreement, thereby affording a much wider scope for the operation of whatever agreement rules the law adopts. Different philosophical theories might even have different implications for the content of those agreement rules — for example, the degree of force needed to make an individual's consent no longer voluntary, or the amount of information needed to make an

37. If the law does not permit a particular background rule to be varied by the parties, then the set of "agreement rules" available to change that background is of course the null set. For a slightly different classification of contract rules, see Ayres & Gertner, *supra* note 3.

individual's consent sufficiently informed. In this article, I am not concerned with the content of contract law's agreement rules, so I will not explore these implications further.

It is less clear that the philosophical literature discussed above has any implications for the content of contract law's background rules. All of the authors discussed above recognize that any real system of promising would have to include some set of rules governing excuses, remedies, and other details of the promisor's obligation.[38] Their purpose, however, was to analyze the practice of promising at a higher level of abstraction. They were interested in the elements all promises had in common, regardless of the action ø that was promised, and largely regardless of the background rules governing such topics as remedies or excuses. Granted, a sufficiently extreme background rule — *e.g.,* one excusing nonperformance whenever the promisor no longer felt like performing — might deprive a so-called "promise" of any binding force at all, thereby eliminating the very aspect of promising that interested these philosophers, and excluding the practice governed by such a rule from the scope of their analysis. Within those limits, though, the rules governing such topics as remedies and excuses could effectively be treated as just a more complete definition of the exact obligation undertaken by the promisor — in other words, as just another aspect of ø. The goal of the writings described above was to explain how and why a promisor could be bound to live up to any ø, regardless of exactly what that particular ø required.

This indifference to the content of ø should not be taken to mean that these authors believed that the appropriate background rules could be determined simply by looking to the explicit content of a party's promise. At a minimum, the parties' specifications would be relevant only if the rule were one the parties were free to vary — a "default rule" in the terminology I have used here. Not all of the writers discussed above are willing to grant the parties that much power over every aspect of their obligation.[39]

Even when a background rule concerns a topic that everyone agrees the parties should be allowed to vary — say, the extent of the warranty in the sale of a used automobile by a private individual — many parties simply will not address that topic in their agreement, so there will be nothing in the agreement's explicit content to resolve this issue. As a result, some method must be found to *interpret* the parties' agreement, to provide rules governing any topic not explicitly settled

38. See, for example, the passage from Rawls quoted in the text *supra* at note 7.

39. *See, e.g.,* Raz, *supra* note 26, at 932.

by the parties. Indeed, creative interpretation is often needed to determine whether there has been any binding promise at all, for even this fundamental question is not always explicitly settled by the parties.[40] While it is perhaps more common to speak of "interpretation" in cases where parties attempt to resolve an issue but do so with insufficient clarity, and to speak of applying default rules in cases where the parties made no attempt to address an issue, the principle is much the same in either case.[41] Both require an outside agency, such as a court, to choose the exact rules defining the parties' obligations where the parties have not unmistakably chosen some rule of their own.

A. *Possible Sources of Law*

The philosophical literature discussed above does not address these issues of interpretation and appropriate default rules. As that literature is concerned with the question of how promises could bind even in the best of circumstances, its focus is implicitly limited to cases where there is no question that a promise has been made, and no difficulty in determining the exact content of the promised action ϕ. That literature may still contribute, however, to the law's resolution of these issues. There are many possible ways of resolving questions of interpretation or background rules, and at least two are perfectly consistent with many of the philosophical positions described above. The first involves a sociological inquiry into the actual practices and customs that exist in any particular community, as a guide to interpreting particular utterances and filling in appropriate background rules. The second involves an appeal to the deeper philosophical values used to justify the institution of promising — that is, values such as social utility or individual freedom or encouraging valuable relationships (depending on the theory of promising employed). Each of these will be discussed below.

1. *Existing Expectations*

The frequent references in the philosophical literature to the "practice" or "institution" of promising could be taken to suggest that the exact scope of any promissory obligation is a matter of sociological fact, to be discovered by careful investigation into the practice of

40. For some famous instances of ambiguity in this respect, see United Steel Workers, Local 1330 v. United States Steel Corp., 492 F. Supp. 1, 5-6 (N.D. Ohio), *affd. in part and vacated in part,* 631 F.2d 1264 (6th Cir. 1980); Embry v. Hargadine-McKittrick Dry Goods Co., 127 Mo. App. 383, 387-92, 105 S.W. 777 (1907).

41. For a more extended discussion of the similarities between interpretation and the selection of default rules, see Goetz & Scott, *supra* note 3, at 264-86.

promising as it exists in the relevant community.[42] For instance, an inquiry into the use of promises in late twentieth-century America might show that promisors were regularly excused whenever performance became commercially impracticable, or it might show that promisors were never excused no matter how difficult performance had become. In either case, the results of that inquiry would define the exact scope of the obligation that any late twentieth-century American had accepted when she made a promise.

Of course, any serious sociological inquiry would very likely identify several different forms of promising, each with different background rules and assumptions, even within a single community.[43] At the very least, it would certainly be *possible* for a society to recognize several different kinds of promises, each with a different set of rules defining the exact scope of the obligation. For example, a society could have one kind of promise that imposes an absolute obligation to perform (in legal terms, one that exposes the promisor to a suit for "specific performance"), another that imposes an obligation to perform or to pay the equivalent in money ("expectation damages"), and a third that imposes an obligation to perform or to make good any losses the promisee may have suffered by relying on the promise ("reliance damages"). The community could also use promises that impose an obligation to perform unless performance became extremely difficult in some unexpected way (in legal terms, promises subject to the defense of commercial impracticability), and promises that permit no such excuse.

The possibility of more than one kind of promise greatly complicates the difficulties involved in interpreting the sociological data about a society's practices. For example, obvious questions arise concerning the number of people who must follow any set of rules for those rules to be accepted as a legally relevant practice. Must an institution be recognized in the community prior to its invocation in any particular transaction, or can any two parties create a custom-made form of promising on the spur of the moment?[44] A related question involves the way we conceive of an individual who appears to be violating the rules of an existing practice: Is she merely an ordinary rule-breaker, or a pathbreaking pioneer in the creation of a new, perhaps

42. I use "sociological" here in its broadest possible sense, to include existing rules of contract law as well as any extra-legal or private promissory practices.

43. As a number of authors have recognized — *e.g.*, J. FINNIS, NATURAL LAW AND NATURAL RIGHTS 308-10 (1980); Jones, *supra* note 16, at 296; Raz, *Promises and Obligations, supra* note 8, at 227-28.

44. The latter position is suggested by Raz, *Promises and Obligations, supra* note 8, at 214-15.

more desirable form of promising? Other difficulties include the problem of conflicting expectations at different levels of generality — for example, if people expect written contracts to be binding, but they also expect goods to be sold at a fair price, what is their expectation regarding the force of a written contract that sets an unfair price? And what of the potential for circularity that arises when people's expectations are themselves affected by existing legal rules?[45]

These problems in inferring morally relevant categories from purely empirical data are well-known, so I will not pursue them here. Instead, for the remainder of this section I will assume that sociologists can identify the set of promises — call them promise$_1$, promise$_2$, . . . promise$_n$ — available to members of any particular society. However, this identification of the relevant choice set does not exhaust the possible uses of sociology. In order to reach a decision about any particular case, the courts must have some method of determining which kind of promise was actually made by the parties to any given transaction.

One can imagine societies in which this second question would be easy to answer — *e.g.,* societies with a system of formal devices by which individuals could signal their choice of institutions. For example, the society might require all binding promises to be signed with a seal (ignoring for the moment the possibility of different kinds of binding promises), while treating any promises not made under seal as nonbinding. In such a society, the problem of interpreting the parties' utterances would deserve the lack of attention it received in the philosophical writings about promises, for it would be, quite literally, nothing but a formality.

The difficulty, of course, is that most societies do not use this method of interpreting parties' utterances, and for good reason. Even when there are only two kinds of promises from which to choose, many writers have commented on the difficulties of expecting all lay people to understand the use of the seal, and the apparent harshness of enforcing one set of rules against parties who clearly intended a different set to apply but who forgot to use the appropriate formality.[46]

45. *See, e.g.,* Feinman, *Critical Approaches to Contract Law,* 30 UCLA L. REV. 829, 837 (1983) ("The law presumably is one factor shaping standards of social and especially commercial behavior. People's 'expectations' and merchants' notions of 'good faith' are to some extent dependent on the positive, public expression of norms by contract law itself."). *But see id.* at 844-45 ("Contemporary and historical research suggests that people generally don't know or don't care much about the rules of contract law and generally carry out their affairs with little regard for them.").

46. *See, e.g.,* Kennedy, *Form and Substance in Private Law Adjudication,* 89 HARV. L. REV. 1685, 1697 (1976).

These difficulties multiply rapidly if there are more than two kinds of promises — that is, more than two permissible sets of background rules — from which to choose. It would be very difficult to design a different seal for each of one hundred possible sets of promissory rules — and even more difficult to expect everybody to remember which seal they should use for each purpose. As a consequence, courts following this approach may have to appeal to sociology not only to identify the set of promises recognized in any particular society, but also to identify the more complex signals by which different kinds of promises are invoked. For example, a survey to determine what people usually mean when they say "I promise" or "I intend" might tell us what kind of commitment people usually have in mind when they make those noises — or, if the inquiry becomes more particularized, when they make those noises in particular contexts. Indeed, this is part of what courts do under current contract law when they inquire into the "reasonable interpretation" of the parties' language.[47]

Notice, though, that this kind of inquiry takes judges far beyond the "small number of basic moral principles" referred to in the introduction to this article. If we must rely on sociological investigation to identify the set of possible background rules, and also to tell us which set of rules applies in any particular case, then sociology is doing virtually all of the work involved in fulfilling the needs of contract law. This alone is enough to defeat the claim referred to in the introduction that the philosophy of promising can by itself yield definite implications for the content of contract law.

2. *Substantive Moral Values*

A more serious objection to this total reliance on sociological data is that it provides no perspective from which to criticize existing promissory practices, or to propose reforms in those practices. One might criticize particular legal rules for not properly conforming to those practices, but there would be no way to criticize the practices themselves. However, sociology is not the only possible source of content for contract law's background rules. An alternative is to look to the substantive values which justify the binding force of promises in the first place (according to one of the philosophical theories discussed earlier), to see if those values have implications for contract law's background rules.

Economic analysis is the most familiar instance of this method of determining the content of contract law's background rules. From an

47. But not the whole part — see *infra* note 57 and accompanying text.

economic perspective, if society is justified in giving individuals the power to make morally binding promises, it is because such promises will, under certain conditions, lead to the most efficient satisfaction of human wants. This notion of efficiency (or some variant of it) can then be used to choose among various possible background rules, by identifying the rule that would contribute most to efficiency. For example, there is an extensive body of literature analyzing different contract remedies to determine which remedies are most efficient in which situations.[48] Economists have also addressed the question of the most efficient rule for excusing promisors who fail to perform because of unexpected difficulties in performance,[49] and the conditions under which individuals' promises should not be treated as binding because of "market failures" that distort the promisor's incentives.[50]

John Rawls provides another example of how background rules might be chosen in order best to serve the substantive values that justify the binding force of promises in the first place. Rawls argued that the binding force of promises is justified if and only if the rules of promising — the background rules, in the terminology used here — are themselves consistent with principles of justice. For Rawls, this meant that they must lead to an equal distribution of all "primary

48. For nontechnical introductions to this literature, see A. POLINSKY, AN INTRODUCTION TO LAW AND ECONOMICS ch. 5 (1983); Craswell, *Contract Remedies, Renegotiation, and the Theory of Efficient Breach*, 61 S. CAL. L. REV. 629 (1988); Kornhauser, *An Introduction to the Economic Analysis of Contract Remedies*, 57 U. COLO. L. REV. 683 (1986). More technical analyses include Cooter, *Unity in Tort, Contract, and Property: The Model of Precaution*, 73 CALIF. L. REV. 1 (1985); Craswell, *Performance, Reliance, and One-Sided Information*, 18 J. LEGAL STUD. 365 (1989); Craswell, *Precontractual Investigation as an Optimal Precaution Problem*, 17 J. LEGAL STUD. 401 (1988); Kornhauser, *Reliance, Reputation, and Breach of Contract*, 26 J.L. & ECON. 691 (1983); Polinsky, *Risk Sharing Through Breach of Contract Remedies*, 12 J. LEGAL STUD. 427 (1983); Rogerson, *Efficient Reliance and Damage Measures for Breach of Contract*, 15 RAND J. ECON. 39 (1984); Shavell, *The Design of Contracts and Remedies for Breach*, 99 Q.J. ECON. 121 (1984); Shavell, *Damage Measures for Breach of Contract*, 11 BELL J. ECON. 466 (1980).

49. *See, e.g.*, Bruce, *An Economic Analysis of the Impossibility Doctrine*, 11 J. LEGAL STUD. 311 (1982); Goldberg, *Impossibility and Related Excuses*, 144 J. INST. & THEOR. ECON. 100 (1988); Joskow, *Commercial Impossibility, the Uranium Market and the* Westinghouse *Case*, 6 J. LEGAL STUD. 119 (1977); Perloff, *The Effects of Breaches of Forward Contracts Due to Unanticipated Price Changes*, 10 J. LEGAL STUD. 221 (1981); Posner & Rosenfield, *Impossibility and Related Doctrines in Contract Law: An Economic Analysis*, 6 J. LEGAL STUD. 83 (1977); Sykes, *Increased Cost of Performance and the Doctrine of Commercial Impracticability*, 19 J. LEGAL STUD. (forthcoming Jan. 1990); White, *Contract Breach and Contract Discharge Due to Impossibility: A Unified Theory*, 17 J. LEGAL STUD. 353 (1988).

50. *E.g.*, Kornhauser, *Unconscionability in Standard Forms*, 64 CALIF. L. REV. 1151 (1976); Ordover & Weiss, *Information and the Law: Evaluating Legal Restrictions on Competitive Contracts*, 71 AM. ECON. REV. PAPERS & PROC. 399 (1981); Rea, *Arm-Breaking, Consumer Credit, and Personal Bankruptcy*, 22 ECON. INQUIRY 188 (1984); Rose-Ackerman, *Inalienability and the Theory of Property Rights*, 85 COLUM. L. REV. 931 (1985); Schwartz & Wilde, *Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis*, 127 U. PA. L. REV. 630 (1979); Schwartz, *A Reexamination of Nonsubstantive Unconscionability*, 63 VA. L. REV. 1053 (1977).

social goods" (liberty, wealth, etc.), except to the extent that an unequal distribution would benefit every member of society.[51] This provides a slightly different criterion for judging possible background rules — although one that will overlap with economic analysis when assessing the ways in which different rules benefit the contracting parties.[52]

Other writers have focused on other substantive values, which could lead them to different conclusions about the proper background rules. For example, John Finnis has argued that the binding force of promises is justified because of the way promises can solve coordination problems.[53] This enables him to endorse whatever background rules are most likely to serve this coordination goal.[54] Still another example can be found in Joseph Raz's suggestion that the binding force of promises is justified by the value of the special relationships which promises create.[55] Though Raz does not develop this theory at any length, it might be possible to decide which of all possible relationships were the most valuable and to adopt the background rules that best facilitated those relationships. For example, if relationships that can be terminated at will are less desirable than those that are more difficult to terminate, that might justify a background rule providing for a relatively large measure of damages for breach of contract.

It is important to realize that the selection of background rules designed to promote the substantive values that justify making promises binding is not necessarily inconsistent with freedom of contract. Under a strong version of this approach — that is, a version arguing that the selected background rules should be *mandatory* — freedom of contract would indeed be restricted.[56] But this approach can also be used in a milder version, endorsing the preferred back-

51. J. RAWLS, A THEORY OF JUSTICE 344-48 (1971). The principles of justice themselves are discussed at more length in *id.* at 54-117.

52. *See id.* at 67-83.

53. J. FINNIS, *supra* note 43, at 298-308. See also the discussion of coordination problems in J. RAWLS, *supra* note 51, at 346-48.

54. *See* J. FINNIS, *supra* note 43, at 320-25 (discussing whether contractual obligations should be viewed as giving the promisor a free election between performing and paying damages).

55. *See* Raz, *Promises and Obligations*, *supra* note 8, at 227-28; *see also supra* note 23 and accompanying text.

56. For various perspectives on arguments that might justify the strong version of one of these theories, see Barnett, *Contract Remedies and Inalienable Rights*, 4 SOC. PHIL. & POLY. Autumn 1986, at 179; Kennedy, *Distributive and Paternalist Motives in Contract and Tort Law, With Special Reference to Compulsory Terms and Unequal Bargaining Power*, 41 MD. L. REV. 563 (1982); Kronman, *Paternalism and the Law of Contracts*, 92 YALE L.J. 763 (1983); Radin, *Market-Inalienability*, 100 HARV. L. REV. 1849 (1987). See also the previously cited economics articles, *supra* note 50.

ground rules merely as default rules or methods of interpretation for those cases where the parties have not specified a preference for some other rule.[57] Any system of law, however committed it may be to the idea of freedom of contract, must have some way of resolving those issues on which the parties' contract is silent or ambiguous. A rebuttable presumption in favor of the rule that best serves some substantive moral value is one way to resolve such cases.

B. *Unhelpful Substantive Values*

The preceding section argued that background rules could sometimes be derived from whatever substantive values justified the binding force of promises in the first place. However, only some of the philosophical theories discussed in section I.B rest on substantive values that are of any help in selecting background rules. In this section, I argue that theories which justify the binding force of promises on the basis of the obligation to tell the truth, or on considerations of individual liberty and autonomy, are of no help at all in such an enterprise. Part III then documents this claim by examining writers who attempt to derive implications for contract law from theories based on individual autonomy or the obligation to tell the truth.

1. *Preventing Misrepresentation*

One of the philosophical theories discussed earlier held that promises derive their binding force from the fact that failing to carry out a promise would falsify the promise as a definite statement about the future, thereby violating the obligation to tell the truth.[58] Assume, for the moment, that this theory is sound. The question is what, if anything, it tells us about the proper background rules of contract law.

With respect to the rules governing implied obligations, such as implied warranties, the misrepresentation theory tells us very little. For example, suppose that someone says, "I promise to give you my car in exchange for $5000," and then delivers a car that doesn't run. The misrepresentation theory says that if the speaker can properly be interpreted as saying "I will definitely give you my car, and it will be in good running condition," then the speaker's failure to do so will make her statement false, thereby violating her obligation to tell the truth. On the other hand, the misrepresentation theory also says that

57. *See, e.g.*, Jacob & Youngs, Inc. v. Kent, 230 N.Y. 239, 242, 129 N.E. 889, 891 (1921) ("From the conclusion that promises may not be treated [as one of the parties had urged] without a sacrifice of justice, the progress is a short one to the conclusion that they may not be so treated without a perversion of intention.").

58. *See supra* section I.B.3.

if the speaker is more properly interpreted as saying "I will definitely give you my car, but I guarantee nothing about its condition," the speaker's failure to deliver a working car will not falsify her earlier statement. In other words, the misrepresentation theory leaves the proper outcome entirely dependent on the proper interpretation of the speaker's promise.

There are, of course, any number of ways to resolve this interpretation question. We could look to sociology to tell us how much responsibility most people expect sellers to assume under these circumstances. Alternatively, we could look to other substantive values to try to resolve the interpretation question. For example, if it is more efficient to put responsibility for such problems on the seller, or if relationships that continue over time are for some reason considered more desirable than relationships where each side's involvement ends as soon as the goods change hands, that might justify interpreting such statements as committing the seller to an implied warranty, at least in the absence of explicit statements to the contrary.

Notice, though, that nothing about the misrepresentation theory tells us which method of interpretation we ought to use. The misrepresentation theory is consistent with interpreting promises in accord with whatever default rule most people already expect, or whatever default rule would be most efficient, or whatever default rule best serves some substantive value other than economic efficiency. While the misrepresentation theory of promising tells us that people must live up to the proper interpretation of their promise, it is equally consistent with any of the ways in which the proper interpretation might be identified.[59] Thus, knowing that contracts should be enforced in order to prevent misrepresentation tells us nothing about which method of interpretation ought to be employed, or anything else relevant to deciding whether there should be an implied warranty in this transaction.

For similar reasons, the misrepresentation theory also tells us little about the background rules governing the proper remedy for nonperformance. If the speaker's promise is interpreted as saying, "I will definitely give you this car, rather than any substitute," then only the delivery of the actual car could avoid falsifying that statement, suggesting that a remedy of specific performance would be most appropriate. But other interpretations are also possible — for example, "I will give you this car, or else give you enough money to let you buy an-

59. For a more extensive discussion of this point, with reference to the law of false advertising, see Craswell, *Interpreting Deceptive Advertising*, 65 B.U. L. REV. 657 (1985).

other car just like it." If this is the proper interpretation to place on the speaker's remarks, ordinary market-value damages would be enough to avoid falsifying her statement. Other interpretations are equally possible — *e.g.*, ". . . or else I'll return your purchase price," ". . . or else I'll make good your reliance expenses," or even ". . . or else I'll pay you *X* dollars to make up for it," where *X* could be any number under the sun.

It might be argued that the misrepresentation theory at least establishes that the seller's failure to deliver the car is wrong, and therefore deserves at least *some* sanction. That is, the misrepresentation theory might seem to point us to punishment theory — to principles of deterrence or retribution, or arguments that the severity of the punishment ought to be proportional to the gravity of the offense — to determine the appropriate sanction for nonperformance. However, even this conclusion depends on an implicit resolution of the interpretation issue discussed in the preceding paragraph. If the seller's statement is best interpreted in one of the ways discussed above — say, "I will definitely give you this car, or else I'll return your purchase price" — then there is nothing wrong with the seller's failure to deliver the car, so long as she is at least willing to return the purchase price. Thus, if the quoted language represents the best interpretation of the seller's statement (according to some theory of what makes for the best interpretation), the question of the proper remedy for nondelivery is settled without any need to appeal to punishment theory. To justify looking to punishment theory to determine the sanction, a misrepresentation theorist would first have to establish that the seller's remarks ought to be interpreted as something other than any of the quoted propositions suggested above.

Admittedly, most promisors probably do not explicitly have in mind anything like those quoted propositions.[60] This may be why it seems inherently correct or natural to treat the question of an appropriate remedy as a question for punishment theory, rather than as a question of interpretation. If most promisors have in mind only something like "I will definitely give you this car," then that certainly *could* be interpreted as an agreement that it would be wrong for the seller to fail to turn over the car, thereby calling on punishment theory to determine the appropriate response. At a minimum, it could be argued that such a construction ought to be adopted as the default rule, thereby shifting to any seller who wanted some other remedy the burden of making a more explicit statement to the contrary.

60. *See* O.W. HOLMES, THE COMMON LAW 237 (1963 ed.) ("when people make contracts, they usually contemplate the performance rather than the breach").

My argument is simply that this construction — the construction that makes the appropriate sanction turn on some theory of punishment — is not in any way entailed by the misrepresentation theory of promising. The value of telling the truth would in no way be compromised by adopting any of the other constructions discussed above, thereby putting the burden of stating otherwise on sellers who object to the chosen construction. Thus, if there is any reason for preferring one possible construction over the others, that reason has to be something more than just the value of telling the truth. It must rest on some belief about what most people already expect, and some argument about why the law ought to fulfill existing expectations; or a belief about what default rule is most efficient, and some argument about why the law should be concerned with efficiency (to cite just two of the possibilities). In other words, the misrepresentation theory settles none of the possible questions about what values the law ought to look to in selecting its background rules.

2. *Individual Autonomy*

Another view of promising justifies the moral force of a promise as a necessary corollary of individual liberty or autonomy.[61] If promises were not binding, it is argued, individual freedom would be unjustifiably restricted, as individuals would be deprived of the freedom to place themselves under a moral obligation respecting their future conduct. While there may be a slight paradox in the notion that freedom must include the freedom to limit one's freedom in the future, advocates of this theory resolve that paradox in favor of allowing individuals to make binding promises.[62]

Autonomy-based theories may well have implications for what I have called "agreement rules," or rules concerning the conditions under which individuals will be allowed to vary the background rules that would otherwise govern their relations. For example, these theorists generally oppose the restrictions on freedom of contract represented by the rule denying enforceability to promises unsupported by consideration, or to promises that are deemed unconscionable.[63] More precisely, they oppose restrictions on the enforceability of promises unless true consent is lacking (*e.g.*, cases of duress), or unless the sub-

61. *E.g.*, C. FRIED, *supra* note 1.

62. *Id.* at 14. *But cf.* 1 M. ROTHBARD, MAN, ECONOMY, AND STATE: A TREATISE ON ECONOMIC PRINCIPLES 153-55 (1962) (offering a libertarian argument that contracts for personal services should never be legally enforceable); *see also infra* note 64.

63. C. FRIED, *supra* note 1, at 28-39, 103-09; Barnett, *supra* note 21, at 313-14.

ject of the promise is not the promisor's to give away.[64] Thus, one necessary part of these theories is a specification of the conditions under which a party's apparent consent will be recognized as valid. As long as these conditions are satisfied, autonomy-based theories hold that any rule or obligation agreed to by the parties should be allowed to govern their relationship.

In cases where the parties have not specified the rule they prefer, however, autonomy-based theories have much less to tell us. In these cases, autonomy-based theories run into the same problem as the misrepresentation theory. Just as any default rule would be consistent with the obligation to tell the truth, any default rule would also be consistent with individual freedom, as long as the parties are allowed to change the rule by appropriate language. Consequently, some other principle must be invoked to decide which of the many possible default rules to adopt. The rule could be chosen by looking to sociological data to determine which rule most parties already expect in various circumstances; it could also be chosen by appealing to some substantive value such as economic efficiency, or Rawls' difference principle, or any other view about what makes some kinds of contractual relationships more valuable than others. Thus, even if those principles have been rejected as valid justifications for the binding force of promises, one or more of them must still be selected to provide the default rules for parties who have not unambiguously specified some other rule in their contract.

The reason that misrepresentation and autonomy-based theories are unhelpful in the selection of default rules is that, of all the philosophical theories discussed earlier in section I.B, these two share the characteristic of being completely content-neutral.[65] They give rea-

64. *Cf.* Barnett, *supra* note 56, at 185-95, 197 (arguing that individuals cannot alienate control over their persons, and thus have no right to provide for the remedy of specific performance in contracts for personal services). Barnett would, however, allow individuals to alienate the disjunctive right to "my personal services or their value in monetary damages, if I subsequently choose not to perform"; he thus would not render contracts for personal services completely unenforceable. *Id.* at 197.

65. A similar concept, "content independence," is employed by Raz, *Authority and Consent*, *supra* note 8, at 114-16 (discussing examples of content-independent reasons for action or belief, including promises); *see also* J. RAZ, PRACTICAL REASON AND NORMS, *supra* note 8, at 70 ("[B]oth [decisions and promises] are content-independent reasons: regardless what you promise or decide to do you have a reason to do it because you have promised or decided.").

Raz argued that theories describing the way in which promises constrain a promisor's subsequent deliberations — that is, theories of the sort discussed earlier in section I.A — had to be content-neutral, or else they could not explain the constraint generated by the promise, as opposed to constraints that might argue in favor of doing ø even if no promise had ever been made. However, Raz in no way suggested that theories explaining the moral force of those constraints (theories of the sort discussed *supra* in section I.B) had to be neutral or indifferent with respect to the content of ø. If anything, his suggestions about the special value of certain kinds of relationships suggest the contrary. Raz, *Promises and Obligations*, *supra* note 8, at 228.

sons why an individual who has promised to do ø thereby incurs some form of obligation to do ø, regardless of how ø is filled in. The reason for this neutrality is understandable: To do anything more requires a theory that would tell people what kinds of promises they ought to make. Unfortunately, the theorists' reluctance to advise individuals as to how they ought to exercise their freedom to fill in the content of ø leaves them equally unable to give legal systems any guidance about how to fill in the content of ø when contracting parties fail to specify their preferred content. As a result, these two theories have nothing to contribute to the selection of default rules governing such important topics as implied terms and conditions, excuses, and remedies for breach.

I should stress that this shortcoming in no way undercuts whatever validity these misrepresentation or autonomy-based theories may have in their original capacity, as justifications for the morally binding force of promises. It simply means that if one of these theories is accepted as the justification for the force of promises, some other theory or theories must then be added to provide a basis for selecting appropriate background rules. This other theory must be a theory that is not neutral between the different ways of filling in the exact scope of the parties' obligation — for example, it must provide some reason for preferring promises with an implied warranty to promises without an implied warranty, or vice versa. In other words, this other theory must rely on more than the value of individual autonomy or the value of telling the truth.

III. TWO EXAMPLES

Part II demonstrated that even if one accepts any of the misrepresentation or autonomy-based theories as justifications for the moral force of promises, there are still important choices to be made concerning the values used to select the law's background rules. Unfortunately, these choices usually receive much less attention from those who write about contract law from the standpoint of misrepresentation or autonomy-based theories. As a result, those authors often end up defending their choice of default rules on an indefensibly ad hoc basis. At other times, they are led to oppose certain default rules unnecessarily, simply because those default rules are supported by certain values that were rivals to the misrepresentation or autonomy-based theory at the level of a justification for the moral force of promises. The final part of this article illustrates the difficulties that such theorists encounter by considering the theories of Charles Fried and Randy Barnett. These two authors have provided the most care-

ful and comprehensive of recent attempts to give contract law a solid philosophical grounding. In addition, each rests his theory of why promises are binding on some version of the value of individual freedom or the obligation to tell the truth.

A. *Charles Fried*

As noted earlier, Fried justifies the obligation to keep a promise primarily by viewing it as a necessary corollary of individual autonomy. "If we decline to take seriously the assumption of an obligation . . . , to that extent we do not take [the promisor] seriously as a person."[66] To be sure, Fried is somewhat ambiguous on this point, for he also emphasizes the injury done to the promisee by the breaking of a promise. For example, he argues that promise-breaking abuses the promisee's trust in the promise, thereby violating the Kantian injunction against treating other people as means rather than as ends.[67] This moves Fried much closer to the misrepresentation theorists: as Fried himself puts it, promise-breaking "is like (but only *like*) lying."[68] But since the autonomy-based and misrepresentation theories are equally unhelpful when it comes to the selection of default rules, an exact classification of Fried's position as between these two theories is unnecessary to my analysis.

1. *Expectation Damages*

On the question of the appropriate remedy for breach, Fried supports the expectation measure of damages, which is designed to give the promisee the same benefits he would have received had the promise been kept. According to Fried, "[i]f I make a promise to you, I should do as I promise; and if I fail to keep my promise, it is fair that I should be made to hand over the equivalent of the promised performance."[69]

A moment's consideration, however, will show that this conclusion cannot be derived solely from the value of individual freedom and autonomy. Fried may well be correct that, in order to give free rein to an

66. C. FRIED, *supra* note 1, at 20-21.

67. *Id.* at 16. I am indebted to John Gardner for pointing out this ambiguity in Fried's position.

68. *Id.* Fried refuses to join the misrepresentation theorists entirely because he interprets the obligation to tell the truth as an obligation not to lie knowingly, thereby imposing no restriction at all on a promisor who at the time of her promise intended to keep the promise. *Id.* at 17. Under the broader version of that obligation endorsed by most misrepresentation theorists, the difference between their theory and Fried's would become very small. *See supra* note 35 and accompanying text.

69. C. FRIED, *supra* note 1, at 17.

individual's autonomy, "[i]t is necessary that I be able to make nonoptional a course of conduct that would otherwise be optional for me."[70] But almost any remedy — reliance damages, punitive damages, specific performance, etc. — makes the promised course of conduct non-optional to some degree, depending on the severity of the threatened penalty. There is surely nothing in the idea of individual autonomy that requires the exact degree of non-optionality provided by the expectation measure. The idea of individual autonomy does suggest that individuals should be allowed to make their conduct nonoptional to any extent they choose, by specifying one of these remedies in their contract. But the law must still select one of these remedies as the default rule, and nothing in the notion of individual autonomy gives any reason for favoring the expectation measure over any of the others.[71]

Fried might, of course, have some other value in mind which explains why the expectation measure is to be preferred (unless the parties specify otherwise) over any of the other possible measures. For example, Fried might believe that the expectation measure promotes economic efficiency, or better satisfies Rawls' difference principle, or would better solve most coordination problems. However, no such argument is made anywhere in his book.

Alternatively, Fried might be appealing to data about people's existing beliefs to justify his preference for the expectation measure. He cannot be relying on existing nonlegal practices, for studies of those practices show that people often do not demand (or offer) expectation damages in cases of unexcused nonperformance.[72] However, Fried might be taking existing *legal* practices as his normative benchmark, for Anglo-American law often does employ the expectation measure of damages. That is, Fried's argument might be that because the law adopts liability for expectation damages as one incident of the obligation of promising, anyone who promises thereby accepts that liability as one of the rules of the game. If this is Fried's argument, though, his theory cannot be what *justifies* the law's choice of the expectation measure. The same argument would work equally well to explain why an individual was obliged to respect any other damage rule the law happened to have adopted.

70. *Id.* at 13.

71. For similar criticisms of Fried's autonomy-based argument for expectation damages, see Epstein, *Beyond Foreseeability: Consequential Damages in the Law of Contract*, 18 J. LEGAL STUD. 105, 106-08 (1989); Farber, Book Review, 66 MINN. L. REV. 561, 564-65 (1982).

72. *E.g.*, Epstein, *supra* note 71, at 112-21; Macaulay, *Non-Contractual Relations in Business: A Preliminary Study*, 28 AM. SOC. REV. 55 (1963).

Fried's only other argument in favor of the expectation measure is really an argument against one of the possible alternatives, the reliance measure of damages. When addressing the possible justifications for the binding force of promises, Fried argues against the theory that promises are binding only because promise-breaking would injure those who relied on the promise. Fried goes on, however, to treat his argument against that justification for the binding force of promises as also being an argument against adopting the reliance measure of damages as a default rule.[73] That is, he seems to assume that the reliance measure of damages could only be justified as a default rule if one first accepted the promisee's reliance as an explanation for the binding force of a promise.

If this is Fried's argument, it rests on a *non sequitur.* Granted, one way of arguing for the reliance measure of damages would be to argue that the reason promises are binding is because promise-breaking injures those who rely on the promise. However, it hardly follows that this is the *only* way that the reliance measure of damages can be supported, or that rejection of the reliance justification for the binding force of promises also entails rejection of the reliance measure of damages. This error becomes even more obvious when it is recalled that we are considering the reliance measure of damages as a default rule, which would still leave the parties free to specify the expectation measure (or any other measure of damages) in its place. Adopting the reliance measure merely as a default rule is in some ways inconsistent with the theory that promises are binding only in order to prevent reliance losses, for a default rule lets the parties agree to other damage rules which would effectively make their promise binding even in the absence of any reliance losses. The notion of reliance damages as a default rule seems much more compatible with the views of someone like Fried, who believes that the binding force of promises derives entirely from the freedom of the individual promisors.

In a nutshell, then, the difficulty with Fried's position on expectation damages is the difficulty identified earlier in Part II of this article. Any damage measure is consistent with the ideal of individual autonomy, as long as it is adopted solely as a default rule, since any default rule expands the promisor's options by making it easier for her to make a certain kind of promise. Fried must therefore invoke some other value in order to decide which of the many damage rules to select as a starting point. Moreover, that value will necessarily be one which Fried rejected as a possible justification for the binding force of promises, for the only value serving that role for Fried — the value of individual autonomy — is equally consistent with all default rules. To choose a default rule, then, Fried has to let one of the other values back into the analysis, and the only question is which one.

73. C. FRIED, *supra* note 1, at 18-19.

2. *Rescission*

Fried encounters a similar problem when he addresses the question of whether a promisee's *only* remedy for nonperformance is to sue for expectation damages, or whether the promisee should also have the option of rescinding the contract and recovering any advance payments. Such an option could be extremely important to the promisee if, for example, the market price of the promised goods had fallen since the time of the contract. In such a case, the promisee would much prefer to get his money back and buy the goods elsewhere (at their new, lower price), rather than merely being allowed to recover from the promisor the market value of any goods that were defective or were not delivered.

Fried initially seems to regard the promisee's right to rescind the contract as being a natural corollary of the binding force of promising. "Parties bind themselves reciprocally. If one party treats himself as not bound, the other may also treat himself as not bound. By breaking his contract, a contractual partner not only opens himself up to claims for damages but releases his opposite number."[74] This argument, though, is subject to the same objections as Fried's argument about the expectation measure. While a system of promising with that default rule would certainly expand a promisor's freedom, so too would an institution of promising with any other rule as its default rule. The quoted passage merely asserts that our system of promising contains rescission as one of its default remedies, without doing anything to justify that rule.

Interestingly, in this case Fried anticipates this very criticism. That is, he acknowledges that it is possible for contracting parties to provide either (a) that nonperformance by one party releases the other party from the contract (in legal terms, making the promises conditional) or (b) that nonperformance by one party does not release the other party from the contract, but only exposes the first party to a suit for damages (in legal terms, making the promises independent). He also notes that such provisions need not be explicitly stated to be effective.[75] In other words, Fried recognizes that there are (at least?) two

74. *Id.* at 117.

75. *Id.* at 118-23.

possible damage rules which parties should be free to invoke, and that the law must somehow decide which rule to treat as invoked in any particular case. As he puts it much earlier in the book, "does your breaking your promise cancel my reciprocal obligation to you or just give me a remedy for my disappointment? There is no obvious a priori reason for one or the other response."[76]

In this case, Fried resolves the dilemma by an explicit appeal to existing expectations, which he reads as supporting a default rule that includes the remedy of rescission. In his words, "[a]ny other outcome would disturb the expectations on which contractual terms *are usually established.*"[77] Unfortunately, Fried says nothing to explain why the expectations of most people in the community should necessarily be dispositive in any individual case. Indeed, at other points in his analysis Fried seems to view the enforcement of community expectations as somehow inconsistent with promissory principles, belonging more to the realm of tort.[78] The problem, of course, is that Fried has to look to *some* outside value to decide whether rescission ought to be accepted as a normal default remedy. In this instance, he "solves" that problem simply by asserting that community expectations ought to govern the matter, without attempting any defense of that position.

3. *Other Default Rules*

Similar difficulties resurface when Fried turns to the rules governing excuse for impracticability, frustration, or mistake. However, Fried takes a somewhat different approach to the selection of the appropriate default rules for these subjects. Fried sees these rules as necessary to fill the "gaps" in the parties' agreement concerning problems which in some sense were unexpected, and for which neither party had agreed to assume responsibility. Because these rules relate to issues outside the scope of the parties' agreement, Fried argues that their justification need not rest on his theory of promises as exercises of individual autonomy. Instead, Fried views these rules as justified by nonpromissory principles such as fault (responsibility for negligently inflicted losses), or what he calls altruism (sharing among members of

76. *Id.* at 46 n.*.

77. *Id.* at 118 (emphasis added). Fried cites no sociological data to support this claim.

78. *Cf. id.* at 4:

> Now tort law typically deals with involuntary transactions — if a punch in the nose, a traffic accident, or a malicious piece of gossip may be called a transaction — so that the role of the community in adjudicating the conflict is particularly prominent. . . . In contrast, so long as we see contractual obligation as based on promise, on obligations that the parties have themselves assumed, the focus of the inquiry is on the will of the parties.

a community).[79] In this context, then, Fried clearly recognizes that he must appeal to substantive, non-content-neutral values.

Fried also believes that many of the rules governing contract formation must depend on values other than individual autonomy. Examples include the common law's "mailbox rule," which provides that an offer cannot be revoked after an acceptance has been posted by the offeree; or the rule that an offer is deemed to lapse (and cannot later be accepted) if it has once been rejected by the offeree. In this area, Fried explicitly recognizes that any number of different default rules are consistent with his theory of promises, and that the offeror should be permitted to specify in her offer whatever rules she wants to apply.[80] Indeed, on the question of which default rule to adopt, he states that "there are no reasons in principle, nothing entailed by the concepts themselves, only considerations of fairness and convenience."[81] In analyzing particular rules, he usually takes the position that the most sensible default rule would be whichever one most offerors would prefer, for that rule saves a majority of offerors the trouble of having to specify a different rule. He then embarks on an analysis of the costs and benefits of each rule to try to determine which rule most offerors would prefer.[82]

In each of these areas, then, Fried is perfectly willing to appeal to some non-content-neutral value such as "fairness and convenience," even though those may not be the values which justify the binding force of promises. Once again, though, Fried never explains why he chooses the particular values he does. For example, why do the rules governing contract formation rest on arguments of "convenience" about the rule that most promisors would prefer, while the rules governing impracticability and mistake rests on principles of "fairness" such as equal sharing? The rules governing impracticability and mistake are always subject to variation by the parties' agreement, so the argument that the law should adopt whatever rule most parties would prefer (in order to save them the trouble of specifying otherwise) would seem just as strong when applied to those issues. Alternatively, if considerations of fairness trump the convenience of the promisor in contracts involving unanticipated risks, why should they not also

79. *Id.* at ch. 5; *see also id.* at ch. 6 (taking a similar approach to the implied obligation of "good faith"). For criticisms of this aspect of Fried's theory, see Atiyah, Book Review, 95 HARV. L. REV. 509, 520-23 (1981), and Reiter, *Good Faith in Contracts,* 17 VAL. U. L. REV. 705, 719-23 (1983).

80. C. FRIED, *supra* note 1, at 47-51.

81. *Id.* at 49.

82. *See id.* at 49 n.* (discussing the rule that an offer lapses once it is rejected by the offeree); *id.* at 52 (discussing the mailbox rule).

trump the convenience of offerors in cases of offers that are delayed in the mail? Without any explicit theory explaining when each of these different values is an appropriate guide, Fried's appeal to different values in different contexts looks more like an *ex post* rationalization of the rules of contract law than a philosophical justification of those rules.

Fried is also on weak ground in explaining his willingness in the areas of impracticability and mistake to make an explicit appeal to values other than individual autonomy, when he was unwilling to allow such appeals in considering the expectation measure of damages. Fried's only justification for looking beyond considerations of individual autonomy when deciding on the rules governing impracticability and mistake is his belief that most parties do not consciously consider the rules they wish to apply to such unexpected contingencies, so enforcing any particular rule can only be justified as a form of involuntary liability.[83] The same is true, however, of parties who make promises without consciously considering what remedies would be available if the promisor fails to perform without an acceptable excuse.[84] If values other than individual freedom and autonomy can be invoked to set the default rules governing impracticability and mistake, it is hard to see why they cannot also be invoked to set the default rules governing remedies for nonperformance.

When all is said and done, then, Fried's theory about what justifies the binding force of promises — the theory which derives promises' force from considerations of individual freedom and autonomy — plays very little role in his derivation of any of the relevant default rules. Unfortunately, Fried's preoccupation with his theory of promising seems to prevent him from developing a coherent theory of the values that *should* play a role in selecting default rules. Sometimes Fried relies on people's existing expectations; sometimes he uses economic arguments; sometimes he rests on principles of "fault" or "altruism"; and sometimes, as in the case of expectation damages, he advances no justification at all. Such a scattershot approach to the selection of default rules does little to advance our understanding of contract law.

B. *Randy E. Barnett*

As noted earlier, Barnett views promises as transferring to the

83. *Id.* at 60.

84. *See supra* note 60. A similar point has been made by Atiyah, *Misrepresentation, Warranty, and Estoppel,* 9 ALBERTA L. REV. 347, 353 (1971).

promisee the promisor's property right in the promised good or service. He therefore treats the obligation to carry out a promise as a specific instance of the more general obligation to respect property rights.[85] While this explanation of the binding force of promises differs from Fried's in some respects, Fried and Barnett are actually very close on the underlying question of why anyone is obliged to respect property rights (in Barnett's case) or promises (in Fried's). Just as Fried defended the obligation to keep a promise on the ground that the power to make binding promises is a necessary corollary of individual liberty, Barnett sees the power to make binding transfers of property as justifiable on essentially libertarian grounds.[86]

Unlike Fried, however, Barnett does not view the obligation to carry out one's promise as binding only when the promisor truly and subjectively agrees to undertake that obligation. Barnett argues that any workable system of property rights must make use of clear signals of entitlement — boundary markers and title recordation in the case of real estate; objective manifestations of consent in the case of contracts — in order to give maximal guidance to those who need to know what their rights are.[87] Thus, in Barnett's view it is not inconsistent with individual liberty to hold individuals liable whenever they manifest their consent to an obligation, even if they subjectively intended to consent to no such thing. According to Barnett, a contrary rule would be inconsistent with the equal liberty of others, who may need to know whether the individual is subject to an obligation or not.[88]

Thus, an important operational difference between Barnett and Fried lies in Barnett's willingness to endorse the objective theory of interpretation as applied to contract law.[89] This frees Barnett from at least one of the difficulties faced by Fried. Fried believed that any background rules pertaining to promissory matters could rest only on

85. *See supra* notes 21-22 and accompanying text.

86. Barnett, *supra* note 21, at 297-99. At least, if one asks why individuals should have the authority to decide whether and when to transfer away their property rights, Barnett's answer is libertarian: Such a right is the best way of "facilitating freedom of human action . . . ." *Id.* at 297. If one asks why freedom of action is itself a desirable thing, his answer sounds closer to economic or utilitarian notions of preference-satisfaction: Liberty is the best way of facilitating individuals' "pursuit of survival and happiness." Barnett, *Pursuing Justice in a Free Society: Part I — Power v. Liberty,* 4 CRIM. JUST. ETHICS 50, 57 (Summer/Fall 1985). Barnett differs from most economists, however, in his willingness to endorse a particular "vision of the good life for men," rather than relying solely on empirical observations about what people happen to prefer. *Id.* at 71 n.49. Since freedom itself is the only aspect of Barnett's vision of the good life that he invokes in his writings about contract law, it seems appropriate to describe his position as resting ultimately on the value of individual freedom.

87. Barnett, *supra* note 21, at 301-07.

88. *Id.* at 305-06.

89. *Id.* at 300-09; *cf.* C. FRIED, *supra* note 1, at 61-67 (criticizing the objective theory of interpretation as inconsistent with individual liberty).

the value of individual autonomy; he therefore had to exclude large areas of contract law (*e.g.,* mistake, impracticability, and offer and acceptance) from the promissory sphere, in order to accept default rules in those areas which were based on other values.[90] Barnett faces no such obstacle to the acceptance of default rules based on other values, for an "objective" interpretation of any given promise will necessarily depend on factors other than the promisor's subjective act of will.

However, Barnett still faces the task of figuring out just which values *should* inform the objective interpretation of any particular action or agreement. Under Barnett's view of promises as marking consent to the transfer of property rights, the rules governing each party's obligations can be described by specifying exactly which rights have been transferred. To use one of Barnett's examples, if a person who agrees to a sale of a car is best viewed as making an unconditional transfer of the right to that car, the buyer would be entitled to specific performance if the seller fails to hand over the property. If the seller is instead viewed as making a conditional transfer of the right to the car *or* the right to damages for nonperformance, the buyer could not sue for specific performance, but only for monetary damages.[91] As Barnett puts it elsewhere, "[m]ost 'real world' contractual disputes involve determining precisely which rights were intended to be transferred by the parties."[92] While this passage could be read as referring to the actual, subjective intentions of the parties, his other writing makes it clear that the only relevant intentions are those indicated by all the objective markers of consent.[93]

Unfortunately, Barnett does not tell us how to decide which rights we should deem transferred by any particular set of objective indicators. In some cases, he looks solely to existing expectations about the obligations normally assumed by parties in similar circumstances. For example, Barnett would usually resolve the issue discussed in the preceding paragraph in favor of the specific performance remedy, treating the parties as agreeing to an absolute transfer of rights (unless the contract specifies otherwise). His principal rationale for this result is that "most people would expect that when the contract is executed [the buyer] has a right to the specified land or car."[94]

90. *See supra* notes 79-84 and accompanying text.

91. Barnett, *supra* note 56, at 195-96.

92. Barnett, *Squaring Undisclosed Agency Law with Contract Theory,* 75 CALIF. L. REV. 1969, 1979 (1987).

93. Barnett, *supra* note 21, at 301-07.

94. Barnett, *supra* note 56, at 195. *But cf. infra* note 95. Barnett also argues that his preferred rule would properly place the burden of arguing against specific performance on the guilty breacher, rather than on the innocent plaintiff. *See* Barnett, *supra* note 56, at 182. However, this argument is clearly incorrect, for we cannot decide whether the seller is guilty of a breach until after we decide how to interpret the transaction. If the transaction is best interpreted as transferring the right to the specified good or to its equivalent in monetary damages, there is nothing "guilty" in the seller paying monetary damages instead of handing over the good.

At other times, Barnett relies on economic arguments about the efficiency of various rules. For example, he suggests that his presumption in favor of specific performance might be reversed in cases where specific performance would be extremely difficult for the seller and a fungible replacement good is easily available to the buyer.[95] On another topic, he argues that an employer who hires an employee but then dismisses him before the job has even begun should normally be treated as having assumed responsibility for any reliance losses the employee suffers — for example, if the employee has given up his prior job or incurred significant moving expenses. Barnett's rationale is a straightforward economic one: employers usually have better information about the risk that the employee will not be needed, so employers will usually be the better risk bearer.[96] At other times, though, Barnett seems to view his theory as not requiring any explicit recourse to economic analysis. For example, he asserts that even if his consent theory is consistent with the dictates of economic efficiency, it is still superior because its outcomes can often be determined "without resorting to an explicit efficiency analysis."[97]

Even more troubling, on other occasions Barnett (like Fried) seems to believe that certain default rules are somehow inherent in the concepts employed by his theory and therefore do not require normative justification of any sort. As an example, consider Barnett's analysis of the undisclosed agency problem which arises when an agent A buys goods from a third party T and turns them over to his undisclosed principal UP, whereupon UP gives A money to pay T for the goods, but A becomes insolvent and T never receives his payment.[98] Barnett concludes that T can sue UP to collect his payment, even though this will make UP pay twice, once to the now-insolvent A and once to T. Barnett's argument is that the agency agreement between UP and A authorized A to transfer to T any of UP's rights. Once A entered into the purchase agreement with T, then, the transfer to T of UP's owner-

95. Barnett, *supra* note 56, at 196 n.59 ("Where these circumstances can be shown to exist. . . it may no longer be safe to presume that sellers would have consented to specific relief."). Barnett would also refuse to allow specific performance in contracts for personal services. *See supra* note 64 and accompanying text. But this is because he believes people should not have the authority to limit their future freedom to this extent.

96. Barnett & Becker, *Beyond Reliance: Promissory Estoppel, Contract Formalities, and Misrepresentations,* 15 HOFSTRA L. REV. 443, 478-80 (1987).

97. Barnett, *supra* note 92, at 1976.

98. *Id.* at 1984.

ship of the purchase price became an accomplished fact, and *UP* therefore had no right to resist *T*'s demand for payment.[99]

While this description of the transaction may be consistent with the rules of agency law, it in no way *justifies* or explains those rules. The problem is that Barnett never justifies his construction of the underlying agency agreement between *UP* and *A*, which he reads as empowering *A* to make an unconditional transfer of *UP*'s rights. At least as a matter of logic, that agency agreement could equally well be interpreted as giving *A* the more limited power to make a conditional transfer of *UP*'s rights (conditional on *A*'s not becoming insolvent), thereby putting the risk of *A*'s insolvency on *T* rather than on *UP*. Either arrangement is perfectly consistent with the principle of giving fair notice to *T*, as long as the legal rule is clear in advance so that *T* knows the degree of risk he is being asked to assume. To put the same point slightly differently, either rule would represent a perfectly "objective" interpretation of the agreement. To argue for one rule over the other, Barnett would have to point to some other reason — efficiency, existing expectations, etc. — to explain why it was better to put the risk of *A*'s insolvency on *UP* rather than on *T*, at least as a presumptive matter. He would then have to explain why the reason he selected (efficiency, or whatever) was the appropriate principle to look to in selecting a default rule.

In discussing a more complex problem of undisclosed agency law, Barnett does suggest an awareness of some of these difficulties. In a footnote, he states:

> Developing a consent theory's approach to construing contractual intent when parties are silent on an issue would require a lengthy and separate treatment. Such an effort would involve, among other topics: (1) a discussion of tacit versus expressed knowledge; (2) the presumption that the parties intended what most similarly situated parties would have intended ex ante, thus putting the onus on a minority of parties to express their dissent from the majority by an express term; and (3) the likely incentive effects of the principles of construction on the bargaining behavior of other parties.[100]

Other topics for discussion could easily be added to this list — for example, (4) the conditions under which transaction costs make it difficult for parties to vary the default rule by an express term to the contrary, and (5) the principle that ought to be used to select the default rule in those cases where whatever rule the law selects is likely to

99. *Id.* at 1984-85.

100. *Id.* at 1986 n.71.

remain in force for most parties.[101]

What Barnett does not seem to realize, however, is how little can be accomplished with respect to the content of the law's background rules until this "lengthy and separate treatment" has been completed. Without a theory of interpretation, the only guidance we are left with in selecting default rules is that the law should take an objective approach to interpretation. But to endorse an objective approach is merely to identify one factor — the secret, subjective intention of either party — which should *not* be used as a reason for preferring one rule over another. It says nothing about which factors *should* be considered. It thus leaves unresolved all the debates concerning the role of efficiency as a goal of contract law, or the extent to which contract law should be shaped by redistributional concerns or other values. Barnett may have rejected those values as inadequate explanations for the binding force of promises, but their role in the selection of default rules remains completely open.

Conclusion

Debates over the question of why promises are binding will no doubt continue to occupy the attention of philosophers and legal scholars. Such debates raise issues that are fundamental to western political thought — issues such as individual freedom versus collective control, economic efficiency versus non-economic values, or (more generally) consequentialist systems of ethics versus deontological ones. Any question that offers such a tempting array of topics will always be the subject of frequent visits by scholars, and the question of why promises are binding is no exception.

My thesis is that debates over the question of why promises are binding do much less than is commonly supposed to settle the role to be played by efficiency, non-economic values, or ethical theories generally in selecting contract law's background rules. More precisely, I have argued that certain answers to the question of why promises are binding do nothing to settle these larger issues. Theories that explain the binding force of promises by pointing to the value of individual freedom, or the obligation to tell the truth, may well be valid answers to the question of why promises are binding. But truth and freedom can usually be served equally well by any background rule, so some other value must be introduced to explain why any one rule ought to be chosen over any other. And when we ask *which* values ought to be

101. For a discussion of these issues from an economic perspective, see the articles cited *supra* in note 3.

introduced for this purpose, we thereby reopen the entire debate about the role of efficiency, non-economic values, and other ethical theories. Even if we have rejected any particular value as an explanation of why promises are binding in the first place, that does not settle the question of whether that value can or should be used in selecting a default rule.

In short, much of the current philosophical debate about the binding force of promises is simply irrelevant to contract law's choice of background rules. Legal philosophers who are interested in the content of contract law should direct their energies to this question, and not solely to the question of why promises ought to be binding.

THE DIVERGENCE OF CONTRACT AND PROMISE

Seana Valentine Shiffrin

TABLE OF CONTENTS

INTRODUCTION 709
I. MAKING ROOM FOR MORAL AGENCY 713
A. Liberalism and the Accommodation of Moral Agents 716
B. What Accommodating Moral Agency Requires 717
II. THE DIVERGENCE OF PROMISE AND CONTRACT 719
A. The Divergence Between Promise and Contract 722
1. Specific Performance and Damages 722
2. Mitigation 724
3. Punitive and Liquidated Damages 726
B. An Objection to the Claim of Divergence 727
III. IS THE DIVERGENCE PROBLEMATIC? 729
A. Efficient Breach 730
B. Distinctively Legal Normative Arguments 733
1. Liquidated Damages and Punitive Damage Agreements 734
2. Gift Promises and Consideration 736
C. Is the Divergence Objectionable? 737
IV. CULTURE AND THE MAINTENANCE OF MORAL CHARACTER 740
A. Poker and Corporate Etiquette 743
B. Individual Versus Corporate Agents 746
V. TOWARD AN ALTERNATIVE CONCEPTION 749

THE DIVERGENCE OF CONTRACT AND PROMISE

Seana Valentine Shiffrin[*]

In U.S. law, a contract is described as a legally enforceable promise. So to make a contract, one must make a promise. The legal norms regulating these promises diverge in substance from the moral norms that apply to them. This divergence raises questions about how the moral agent is to navigate both the legal and moral systems. This Article provides a new framework to evaluate the divergence between legal norms and moral norms generally and applies it to the case of contracts and promises. It introduces and defends an approach to the relationship between morality and law that adopts the perspective of moral agents subject to both sets of norms and argues that the law should accommodate the needs of moral agency. Although the law should not aim to enforce interpersonal morality as such, the law's content should be compatible with the conditions necessary for moral agency to flourish. Some aspects of contract not only fail to support the morally decent person, but also contribute to a legal and social culture that is difficult for the morally decent person to accept. Indeed, U.S. contract law may sometimes make it harder for the morally decent person to behave decently.

INTRODUCTION

In U.S. law, a contract is described as a legally enforceable promise. So to make a contract, one must make a promise. One is thereby simultaneously subject to two sets of norms — legal and moral. As I argue, the legal norms regulating these promises diverge in substance from the moral norms that apply to them. This divergence raises questions about how the moral agent is to navigate both the legal and moral systems. In this Article, I provide a new framework to evaluate the divergence between legal norms and moral norms generally and apply it to the case of contracts and promises.

By claiming that contract diverges from promise, I mean that although the legal doctrines of contract associate legal obligations with morally binding promises, the contents of the legal obligations and the legal significance of their breach do not correspond to the moral obligations and the moral significance of their breach.[1] For instance, the

[*] Professor of Philosophy and Law, University of California, Los Angeles. I am grateful for help and critical comments from Iman Anabtawi, Sarah Coolidge, Richard Craswell, Meir Dan-Cohen, Ronald Dworkin, Melvin Eisenberg, Barbara Fried, James Gordley, Mark Greenberg, Tom Grey, Jeffrey Helmreich, Barbara Herman, Doug Kysar, Stephen Munzer, Liam Murphy, Thomas Nagel, Melanie Phillips, Todd Rakoff, Jennifer Roche, William Rubenstein, Steven Shiffrin, and audience members at the Analytical Legal Philosophy Conference, Boalt Hall, the Harvard Law School Faculty Colloquium, the NYU Law and Philosophy Colloquium, the September Group, and the Stanford Legal Studies Colloquium. This Article began as the Kadish Lecture at Boalt Hall and I am especially grateful to Sandy Kadish for its instigation.

[1] Although this Article discusses those aspects of contract that diverge from promissory norms, other parts of contract law depend fairly explicitly upon a variety of moral judgments and

moral rules of promise typically require that one keep a unilateral promise, even if nothing is received in exchange. By contrast, contract law only regards as enforceable promises that are exchanged for something or on which the promisee has reasonably relied to her detriment. When breach occurs, the legal doctrine of mitigation, unlike morality, places the burden on the promisee to make positive efforts to find alternative providers instead of presumptively locating that burden fully on the breaching promisor. Morality classifies intentional promissory breach as a wrong that, in addition to requiring compensation, may merit punitive reactions, albeit sometimes minor ones; these may include proportionate expressions of reprobation, distrust, and self-inflicted reproofs, such as guilt. Contract law's stance on the wrongfulness of promissory breach is equivocal at best, manifested most clearly by its general prohibition of punitive damages.

To analyze this substantive divergence between legal and moral norms, I introduce and defend an approach to the relationship between morality and law that adopts the perspective of moral agents subject to both sets of norms and argue that the law should accommodate the needs of moral agency. Although the law should not aim to enforce interpersonal morality as such, the law's content should be compatible with the conditions necessary for moral agency to flourish. Some aspects of U.S. contract law not only fail to support the morally decent person, but also contribute to a legal and social culture that is difficult for the morally decent person to accept. Indeed, U.S. contract law may sometimes make it harder for the morally decent person to behave decently.

For some, the divergence between contract and morality calls for a direct condemnation of contract law's content. If contract law's business is to enforce promises, its structure, as a whole, should reflect the moral structure of promises.[2] Others regard the divergence between

concepts, not to mention the concept of a promise itself. For instance, the past consideration doctrine contains an exception for cases in which there is an independent moral obligation to do what one has promised. *See* E. ALLAN FARNSWORTH, CONTRACTS §§ 2.7–2.8, at 56–63 (4th ed. 2004). Some cases and doctrines involve appeal to moral concepts of fairness or reasonableness. Not all of these cases and doctrines can be recast more narrowly as judgments about what *justice* requires. Aspects of the immorality doctrine, for instance, may not be reducible to efforts to enforce the spirit or the letter of other aspects of public policy. This partial convergence between morality and contract would call for justification if one took the position that contract and promise are entirely separate domains.

[2] *See, e.g.*, Peter Linzer, *On the Amorality of Contract Remedies — Efficacy, Equity, and the Second* Restatement, 81 COLUM. L. REV. 111 (1981); Frank Menetrez, *Consequentialism, Promissory Obligation, and the Theory of Efficient Breach*, 47 UCLA L. REV. 859, 879–80 (2000). Professor Charles Fried seems to take this position in *Contract As Promise*. *See* CHARLES FRIED, CONTRACT AS PROMISE 1–3, 17–21 (1981); *see also id.* at 17 ("The moralist of duty thus posits a general obligation to keep promises, of which the obligation of contract will be only a special case — that special case in which certain promises have attained legal as well as moral force."). At

contract norms and moral norms as not significant per se, whether because the law, especially private law, should not directly enforce morality as such, or because promise and contract occupy different realms with independent purposes: promise establishes rules for formalizing trust in interpersonal interactions; contract establishes rules that help to enable a flourishing system of economic cooperation for mutual advantage.[3]

My own view does not fit neatly into either of these camps, which I see as marking two poles of a broader, but artificial, dichotomy. This dichotomy seems to be the product of a set of familiar but overly general, overly blunt questions, such as whether law should reflect interpersonal morality and, in particular, whether contract should mirror the moral rules of promising. These questions frame the issues in ways that obscure an important position about the relation between legal and moral norms that maintains that the law should accommodate moral agency, but neither directly reflect nor entirely ignore interpersonal morality. The law must be attentive to the full range of normative positions because law represents a special form of normative cooperative activity. Yet, because law is a cooperative activity of mutual governance that takes institutional form, its normative values and

points, though, Professor Fried's position is more qualified and ambiguous. It may well be compatible with the ideas I develop below.

Professor Fried's substantive criticisms of contract's moral content, however, are limited to the consideration doctrine. Interestingly, Professor Fried does not discuss punitive damages. He endorses the mitigation doctrine, characterizing it as an altruistic duty toward the promisor that is without cost to the promisee. *See id.* at 131. Oddly, given his methodology, he does not consider whether the norms of promising in fact include this duty. *See* LOUIS KAPLOW & STEVEN SHAVELL, FAIRNESS VERSUS WELFARE 161 n.18 (2002) (mentioning Professor Fried's puzzling endorsement of expectation damages rather than specific performance); DORI KIMEL, FROM PROMISE TO CONTRACT 110–11 (2003) (discussing Professor Fried's treatment of mitigation).

[3] *See* FRIED, *supra* note 2, at 2–3 (citing PATRICK ATIYAH, THE RISE AND FALL OF FREEDOM OF CONTRACT (1979); LAWRENCE FRIEDMAN, CONTRACT LAW IN AMERICA (1965); GRANT GILMORE, THE DEATH OF CONTRACT (1974); MORTON J. HORWITZ, THE TRANSFORMATION OF AMERICAN LAW, 1780–1860, at 160–210 (1977); Duncan Kennedy, *Form and Substance in Private Law Adjudication*, 89 HARV. L. REV. 1685 (1976); Duncan Kennedy, *The Structure of Blackstone's Commentaries*, 28 BUFF. L. REV. 205, 356 (1979); Anthony T. Kronman, *Contract Law and Distributive Justice*, 89 YALE L.J. 472 (1980); Ian R. Macneil, *The Many Futures of Contracts*, 47 S. CAL. L. REV. 691 (1974)); *see also* KAPLOW & SHAVELL, *supra* note 2, at 162–65, 182–85, 190–97 (arguing that promise-keeping notions conflict with a welfare-maximizing view of contract and advocating the latter); Richard Craswell, *Contract Law, Default Rules, and the Philosophy of Promising*, 88 MICH. L. REV. 489 (1989) [hereinafter Craswell, *Contract Law*] (arguing that theories of promise have little relevance for the "background rules" of contract); Richard Craswell, *Two Economic Theories of Enforcing Promises*, *in* THE THEORY OF CONTRACT LAW 19 (Peter Benson ed., 2001) [hereinafter Craswell, *Two Economic Theories*]; Anne de Moor, *Are Contracts Promises?*, *in* OXFORD ESSAYS IN JURISPRUDENCE 103 (John Eekelaar & John Bell eds., 3d series 1987) (casting doubt on the equation of contract with promise).

principles may well be distinct from, though informed by, those comprising interpersonal morality.[4]

This Article approaches these topics by exploring the demands and tensions to which contract law, in particular, subjects moral agents. It starts from the more general premise that law must be made compatible with the conditions for moral agency to flourish — both because of the intrinsic importance of moral agency to the person and because a just political and legal culture depends upon a social culture in which moral agency thrives. The content and normative justifications of a legal practice — at least one that is pervasive and involves simultaneous participation in a moral relationship or practice — should be capable of being known and accepted by a self-consciously moral agent.[5] Legal rules must be constructed and justified in ways that take into account the fact that law embodies a system of rules and practices that moral agents inhabit, enforce, and are subject to alongside other aspects of their lives, especially their moral agency. Although I provide some motivation for these premises from a liberal perspective, for the most part I take them as starting points.

Part I expands briefly on these premises and on the distinctive liberal approach they offer to problems concerning the intersection of law and morality. The remainder of the Article applies this approach to the specific problems raised by the divergence between contract and promise. The overarching aim is to explore how an accommodationist approach would frame some issues in contract and to identify some new questions it would raise. This Article does not champion particular positions about the proper content of contract law, although toward the end it gestures at an alternative theoretical conception of contract informed by an accommodationist approach.

Part II argues that contract and promise do indeed diverge. Part III argues that the divergence is problematic. Although the divergence may not yield contradictory requirements, some prominent justifications for the divergent aspects of contract could not be known and accepted by moral agents. Further, the divergence raises the concern that the culture created by contract law and its justifications might

[4] *See* Barbara Herman, *Reasoning to Obligation*, 49 INQUIRY 44, 60 (2006) (observing that coercion may be morally permissible and even required within a legal institutional context even though it would not be justified in interpersonal contexts); Seana Valentine Shiffrin, *Speech, Death, and Double Effect*, 78 N.Y.U. L. REV. 1135, 1181–84 (2003) (discussing the general point in the context of the First Amendment and rules that govern intentions).

[5] In a prior work, I used the perspective of our collective role as moral agents engaged in enforcement to justify the unconscionability doctrine. *See* Seana Valentine Shiffrin, *Paternalism, Unconscionability Doctrine, and Accommodation*, 29 PHIL. & PUB. AFF. 205 (2000). Such arguments might extend to some applications of the immorality doctrine as well. In this Article, I consider our role as subjects of contractual rules, albeit subjects who have a role in the authorship of the rules to which we are subject.

make it more difficult to nurture and sustain moral agency, in particular the virtues associated with fidelity. Part IV rejects the suggestion that these difficulties could be bypassed by disentangling promise and contract through an explicit reconception of contract law as a normative system utterly distinct from promises. Contract law must at least be constrained by some of the needs and rationales of the system of promising. Part V closes with preliminary thoughts about how contract theory could be continuous with (rather than merely cabined by) these constraints.

I. MAKING ROOM FOR MORAL AGENCY

Two standard strains of argument address the relationship between contract and promise. Reflective approaches take interpersonal morality as a template for legal rules, sometimes implicitly. Such approaches operate as though the law should reflect everyday moral judgments whenever possible, whether because this is the nature of law or because, as a matter of political philosophy, it is what law should aim to do. A reflective approach may be particularly tempting in contract law because contracts and promises are so closely intertwined.

By contrast, separatist approaches treat law and morality generally, or contract and promise in particular, as independent domains. Some separatists regard reflective approaches as illiberal. For normative, political reasons, they regard it as inappropriate for the law to incorporate or enforce the rules of interpersonal morality as such. Other separatists have different, positive reasons for treating the domains as distinct. They believe that law, or perhaps specifically contract, has its own goals and purposes, such as establishing the foundation for a maximally efficient system of exchange. Their pursuit does not require engagement with other moral concerns. So, it is not the business of contract law per se to reflect or respond to the norms of interpersonal morality. Separatists acknowledge that individuals may be subject to moral demands that regulate their behavior, but separatists nonetheless maintain that compliance with these demands is each individual's responsibility. Since contract law's purpose is not to enforce interpersonal morality or promises as such, there is no special worry associated with contract law's divergence from promise. Participants in contract, as in other legal domains, may have to constrain their pursuit of its goals to accommodate other personal and social values, but that represents a flaw neither in the separatist conception of contract nor in its divergence from promise.

Both approaches harbor some elements of truth and so neither seems correct. I subscribe to an intermediate position that advocates accommodating moral agency but not enforcing morality as such. I agree with separatists that there is no direct and reliable route from the content of interpersonal morality to the appropriate content of the

corresponding area of law. Legal domains may pursue normative purposes and principles of their own that are not straightforwardly derived from interpersonal morality. Furthermore, the standard liberal concerns about direct enforcement of morality have some traction.

It does not follow, however, that legal principles in these domains should be entirely insensitive to or divorced from the demands of interpersonal morality. Such insensitivity is not dictated by a commitment to liberal principles. Liberalism does not require that the theory of justice and the law must be predicated on a model that works equally well for amoral and moral agents. Quite the contrary. John Rawls's theory, for example, explicitly and plausibly presupposes that agents have developed moral capacities that underlie their capacity for a sense of justice.[6] These moral capacities are not so fine-tuned that they consist of only those abilities necessary to understand and comply with the terms of justice narrowly understood. Rather, they depend on a broader, fuller moral personality.

Indeed, mastery and appreciation of promissory norms must figure among these requisite moral capacities, however broadly or narrowly construed.[7] Absent a culture of general mastery and appreciation of promissory norms and the moral habits and sensitivities that accompany them, I doubt that a large-scale, just social system could thrive and that its legal system could elicit general patterns of voluntary obedience. Further, I doubt that, absent a strong promissory culture, the individual relationships that give rise to and sustain moral agency and relationships of equality could flourish. Whether or not all norms and virtues of moral agency must be accommodated by the legal system, those associated with promissory norms seem quite central to the possibility of a flourishing system of justice.[8]

---

[6] *See* JOHN RAWLS, A THEORY OF JUSTICE 395–587 (1971).

[7] *See* Seana Valentine Shiffrin, Promising, Intimate Relationships, and Conventionalism 42–52 (Dec. 1, 2006) (unpublished manuscript, on file with the Harvard Law School Library) (arguing that promisors must have the ability to bind themselves with a promise "if they are to have the ability to conduct relationships of adequate moral character"); *see also* H.L.A. HART, THE CONCEPT OF LAW 197 (2d ed. 1994) (recognizing the social necessity of rules respecting promises); RAWLS, *supra* note 6, at 346–48 (noting the obligation, derived from the "principle of fairness," to comply with promises). This is not to endorse the view that the duty to obey rests on a promise or a contract.

[8] Some reports about Tongan and Iranian culture may complicate claims about the universality of promising, but it is difficult to determine exactly what they show. *See* Fred Korn & Shulamit R. Decktor Korn, *Where People Don't Promise*, 93 ETHICS 445 (1983) (discussing the absence of promising in Tonga island culture); Michael Slackman, *The Fine Art of Hiding What You Mean To Say*, N.Y. TIMES, Aug. 6, 2006, § 4 (Week in Review), at 5 (discussing the prevalence of what appear to be insincere promises in Iranian culture). Some of this evidence could be interpreted to show that in some cultures, it is hard to tell when sincere commitments are expressed, but not that promises as such have little or no significance. Further, the purported departures from the culture of promising take place in contexts of hierarchical, unequal structures; thus, they may not serve as counterevidence of the importance of promising in maintaining social relations of equality amid local conditions of vulnerability. Tongan society is so pervasively hierarchical that much of its social interaction and dialogue is colored by status differences. *See, e.g.*, Kerry E. James, *Tonga's Pro-Democracy Movement*, 67 PAC. AFF. 242, 243 (1994) (noting the Tongan "traditions of rank and hierarchy"); Adrienne L. Kaeppler, *Poetics and Politics of Tongan Laments and Eulogies*, 20 AM. ETHNOLOGIST 474, 476 (1993) (drawing on Tongan rituals and pronouncements to show that "hierarchical principles of status and rank pervade life and death"); Adrienne L. Kaeppler, *Rank in Tonga*, 10 ETHNOLOGY 174, 174, 177, 188, 191 (1971) (arguing that the inequality inherent in Tongan society remains in force and has withstood the small movement for democratic reform); Kerry E. James, *Pacific Islands Stakeholder Participation in Development: Tonga* 17 (World Bank, Pac. Islands Discussion Paper Series No. 4, 1998) (describing Tongan society as traditional and hierarchical).

Similarly, Slackman describes the Iranian practice of *ta'arof*, a form of polite communication generally regarded as insincere, as developing from the need for subterfuge, against occupiers for example. *See* Slackman, *supra*. While perhaps not as all-encompassingly hierarchical as Tongan society, the context in which *ta'arof* occurs reflects unequal status relationships. *See, e.g.*, WILLIAM O. BEEMAN, THE "GREAT SATAN" VS. THE "MAD MULLAHS": HOW THE UNITED STATES AND IRAN DEMONIZE EACH OTHER 43 (2005) (tracing U.S. misunderstanding of Iran to Americans' failure to appreciate, among other things, that Iranian communication "tends toward hierarchical skewing"); Anne H. Betteridge, *Gift Exchange in Iran: The Locus of Self-Identity in Social Interaction*, 58 ANTHROPOLOGICAL Q. 190, 191 (1985) (arguing that *ta'arof* reflects strategic rules for interaction among people of differing status, as contrasted with "intimates"); Zohreh R. Eslami, *Invitations in Persian and English: Ostensible or Genuine?*, 2 INTERCULTURAL PRAGMATICS 453, 456 (2005) (finding that communication involving *ta'arof* reflects the communicators' consciousness of their status differences); Laurence D. Loeb, *Prestige and Piety in the Iranian Synagogue*, 51 ANTHROPOLOGICAL Q. 155, 156 (1978) (pointing out that *ta'arof* was used "by the [Persian] elite to reinforce rank differentiation"). Professor Loeb shows that a similar consciousness of rank guides the practice of *ta'arof* in a contemporary Iranian synagogue. Loeb, *supra*, at 157.

When the directives of law and morality regulate the same phenomena, moral agents have to negotiate two distinct sets of norms. The rules of contract and promise present a salient instance of this navigation problem. Especially because there are moral duties to obey the law, legal rules should be sensitive to the demands placed on moral agents so that law-abiding moral agents do not, as a regular matter, face substantial burdens on the development and expression of moral agency.[9] Respecting this constraint yields an intermediate stance between the position that law should promote moral behavior for its own sake and the position that interpersonal morality should be irrelevant to the form and justification of law. To wit, even if enforcing interpersonal morality is not the proper direct aim of law, the requirements of interpersonal morality may appropriately influence legal content and legal justifications to make adequate room for the development and expression of moral agency. The influence that is exerted, however, need not result in efforts to mirror interpersonal morality closely.

---

[9] I assume that the rules of morality have much substantive content independent of and prior to the law. That is, their content is not fully determined by law's contents and those moral principles requiring obedience to law.

A. Liberalism and the Accommodation of Moral Agents

These reasons to reject separatist views do not commit us either to a version of reflectivism or to some other illiberal approach. We can be sensitive to the conditions for supporting moral agency and, for that reason, fashion law to be responsive to the content of interpersonal morality without running afoul of liberal strictures that counsel against enforcement of morality or that declare the priority of the right over the good.[10] Two main commitments underlie these strictures. First, the theory of justice must be articulated and defended without relying on any particular comprehensive theory of the good. Its contours are not subject to objection on the mere grounds that they fail to promote or reflect the commitments of any such theory. Second, agents have a primary commitment to abide by the requirements of justice, a commitment that overrides conflicts presented by agents' specific conceptions of the good.

These commitments do not preclude efforts to ensure law's compatibility with the conditions for moral agency. The idea that the right is prior to the good does not entail that the law's content should never be sensitive to the norms of interpersonal morality. The requirements of justice do not fully determine all aspects of law. Contract provides a salient example. The principles of right may require a system of distributive justice, which may in turn require a system of contract. Any adequate contract law may need to have certain features, but all of its particulars may not be fully determined by the principles of justice narrowly construed. For example, considerations of justice may not decisively settle whether to adopt the mailbox rule. Settling the specifics of contract law may depend, in part, on other features of the culture and the society, including components of interpersonal morality. For instance, *if* contract law's aim were to protect against harm suffered from breach of promise, measured in terms of reasonable reliance, what counts as a reasonable form of reliance might depend on the cultural context and the degree to which easy trust is encouraged; the degree to which trust is encouraged might, in turn, be a matter settled partly by the norms of morality and not merely by cultural customs.

To be sure, in many circumstances, liberal principles of justice may preclude the direct implementation of moral commitments for their own sake or direct appeal to them as a rationale for legal decisions. Such principles prohibit both direct and indirect state pressure to engage in activities whose value depends on authentic and voluntary

[10] *See, e.g.*, JOHN RAWLS, POLITICAL LIBERALISM 173–211 (1993); JOHN RAWLS, *The Priority of the Right and Ideas of the Good*, *in* COLLECTED PAPERS 449 (Samuel Freeman ed., 1999).

participation, such as which private relationships to pursue or which ideas to express. Think of the speech and religion protections. Further, to invoke a familiar Rawlsian theme, liberal principles prohibit state endorsement of positions or values, social disagreement about which creates the very need for a theory and system of justice, but reliance on which is not necessary for the system to fulfill its purposes.

Not all components of moral agency or principles of morality fall into these special categories. In particular, the rules of promissory commitment generally do not demand authentic endorsement or performance for their value, nor are they controversial in the ways that create the very need for a system of justice.[11]

I advocate an intermediate position — that law's contents must be structured to make room for moral agents. It calls for the accommodation of moral agents, although the envisioned accommodation differs from some other common forms and connotations of accommodation.[12] Often, "accommodation" refers to adjustments or exceptions to otherwise valid laws that are made for unusual agents or outliers. The sort of accommodation defended here, by contrast, is more foundational. It would inform the guidelines for the shape of the general law and would extend to all moral agents. It would not merely and weakly exempt some agents from legal requirements to permit them to engage in certain behaviors or ways of life. This version of accommodation contains the stronger, positive requirement that the legal system's rules and justifications should be acceptable to moral agents without disrupting their moral agency.

B. What Accommodating Moral Agency Requires

I have been arguing that the legal system should be fashioned, justified, and interpreted to accommodate the opportunity for the governed to lead a full and coherently structured moral life. What does this commitment entail in the case of contract? I start from the following premise: when a legal practice is pervasive and involves simultaneous participation in a moral relationship or practice, the content and normative justification for the legal practice must be acceptable to a reasonable moral agent with a coherent, stable, and unified personality. Law's justification should not depend upon its being opaque or

[11] Performance of some intrafamilial promises may be an exception, and this may in part (though only in part) explain the resistance to enforcing intrafamilial promises. *See, e.g.*, KIMEL, *supra* note 2, at 72–87 (discussing the inadequacy of the contract framework for pursuing the value of personal relationships); *see also infra* section II.A.2, pp. 724–26 (discussing mitigation); *infra* section III.B.2, pp. 736–37 (discussing gift promises).

[12] I discuss some more familiar forms of accommodation and their role within liberalism in Seana Valentine Shiffrin, *Egalitarianism, Choice-Sensitivity, and Accommodation*, *in* REASON AND VALUE: THEMES FROM THE MORAL PHILOSOPHY OF JOSEPH RAZ 270 (R. Jay Wallace et al. eds., 2004), and Shiffrin, *supra* note 5.

obscure or upon the ignorance, amorality, or split personality of the citizens it governs.[13] From this basic premise follow three more specific principles that regulate the interrelation between moral norms and legal norms, at least in those pervasive, regular contexts that involve the simultaneous participation of moral agents in parallel legal and moral relationships or practices.

First, what legal rules directly require agents to do or to refrain from doing should not, as a general matter, be inconsistent with leading a life of at least minimal moral virtue. "Minimal moral virtue" should be understood in a way that does not presuppose any particular comprehensive conception of the good or ideal of virtue.

Second, the law and its rationale should be transparent and accessible to the moral agent. Moreover, their acceptance by the agent should be compatible with her developing and maintaining moral virtue. Although knowledge of the justifications of law is not required or expected of every citizen, understanding the law's rationale should not present a conflict for the interested citizen qua moral agent. This is not merely because the agent is subject to the law and that to which she is subject should be justifiable to her. Within a democratic society, the law should be understood as ours — as authored by us and as the expression of our joint social voice.

This second principle governs both the reasons that actually motivate government agents to impose and enforce divergent rules, as well as the strongest available justifications for the divergence. Examination of the latter reveals whether the divergence is intrinsically problematic. This task is the focus of this Article. Nonetheless, it is possible for there to be an adequate theoretical justification for a divergent rule that is not actually the basis of a court's or a legislature's adoption of that rule. A court or legislature might be guided by poor reasons even when better reasons could be provided for the position. In such a case, we might say that the particular instance of imposition was impermissible because its rationale was not acceptable to a moral agent, even if the rule itself is not intrinsically problematic.[14] In addition, adoption of an intrinsically permissible rule for an impermissible reason may be problematic because even the implicit endorsement of unacceptable reasons[15] may have a corrosive effect on the moral culture and thereby implicate the next principle.

---

[13] I use "citizens" as a shorthand for those regularly subject to a legal regime's requirements.

[14] I note, but do not address here, the further question of whether a regime that showed *convergence* between legal and moral rules, but whose legal justifications were unacceptable to moral agents, would also be suspect. I believe it would be, but that position requires greater argument than there is room for in this Article.

[15] As is well known, there are concerns about how we can coherently refer to the reasons that motivated a legislature, given that it may pass legislation without discussion of, much less consensus on, reasons for that legislation. I put these issues aside here, although a fuller account of legislative rationale must face them, as some have attempted to do. *See, e.g.*, Elizabeth S. Anderson & Richard H. Pildes, *Expressive Theories of Law: A General Restatement*, 148 U. PA. L. REV. 1503, 1520–27 (2000) (asserting that the expressive meaning of collective action is not only a matter of legislative intent, but also a function of public understanding); Elena Kagan, *Private Speech, Public Purpose: The Role of Governmental Motive in First Amendment Doctrine*, 63 U. CHI. L. REV. 413, 438–42 (1996) (arguing that governmental motive could be determined by looking at the legislation produced).

Third, the culture and practices facilitated by law should be compatible with a culture that supports morally virtuous character. Even supposing that law is not responsible for and should not aim to enforce virtuous character and interpersonal moral norms, the legal system should not be incompatible with or present serious obstacles to leading a decent moral life. A principled requirement that the law facilitate a culture that is compatible with moral virtue need not go so far as to enforce moral virtue. In some circumstances, this goal may better be realized by doing quite the opposite. One may facilitate moral virtue by affording opportunities to be virtuous and by refraining from offering strong incentives or encouragements to misbehave; direct enforcement of virtue for its own sake, in some contexts, can be counterproductive — particularly when it is a necessary aspect of the virtuous conduct that it be voluntary and that it be evident to others that it is voluntarily performed.[16]

The remainder of my discussion investigates whether the divergence of contract from promise can satisfy these principles. The next two Parts argue that although contract law does not violate the first principle by issuing directives that contradict moral requirements, it may violate the second and third principles. Defenses of the rules of contract law often invoke rationales the acceptance of which is incompatible with maintaining one's moral convictions. Further, the rules themselves may contribute to a culture that challenges moral agency.

II. THE DIVERGENCE OF PROMISE AND CONTRACT

I now turn to the particulars of the divergence between contract and promise. I focus specifically on the ways in which contract law expects less of the promisor and more of the promisee than morality does.[17] Chiefly, I examine the treatment of remedies in contract and promise, although I occasionally draw on other helpful examples.

---

[16] For a discussion of the claim that factors of this sort support the consideration rule, see *infra* section III.B.2, pp. 736–37.

[17] In some contexts, however, contract may be more demanding of the promisor than morality. This is certainly the popular notion of contract among laypeople. In close interpersonal relationships, there may be more room for interpretative flexibility than in contract. The law must come to a definitive conclusion about what a clause means; sometimes this serves the promisee's interests, as when the promisor is deemed to bear the risk of ambiguity — perhaps because he is the drafter or because he knows more about the industry. *See* RESTATEMENT (SECOND) OF CON-

In this Part and the two Parts following, the argument proceeds in three steps that correspond to the three principles for accommodating moral agency just articulated. First, contract and promise diverge in some significant ways, although not by directly generating inconsistent directives. Second, some of the standard arguments for the doctrines' divergence are exactly the sort of justifications that a virtuous agent could not accept. Third, even though some reasons for the divergence may be acceptable to a virtuous agent, the divergence itself may risk another difficulty by contributing to a culture that may be in tension with the conditions for the maintenance of moral character.

Before launching this argument, a few further assumptions about contract and promise should be made explicit. With respect to promising, I will assume, but not defend, that there are definite norms of promising that all moral agents are required to respect.[18] Generally, however, all that is required to raise the sorts of questions contemplated in this Article is the weaker assumption that there are *some* definite promissory norms — whether or not they are exactly as I describe — that bind moral agents within an aspirationally democratic, large-scale society such as ours.[19]

Of course, the moral contours of the promissory terrain are not always evident on the ground. In addition, there is a range of commit-

TRACTS § 201(2) (1979). By contrast, moral partners may be obliged to acknowledge directly the ambiguity of the meaning of a promise and strike a compromise given two competing reasonable understandings. Further, as Professor Bernard Williams notes, in many informal contexts promissory parties should adjust their understandings of one another when compliance becomes more difficult than anticipated, especially when unforeseen circumstances arise. BERNARD WILLIAMS, TRUTH AND TRUTHFULNESS 112 (2002). By contrast, contract law adopts stricter rules, placing the burden on the party who reasonably bears the risk of unforeseen circumstances. *See, e.g.*, RESTATEMENT (SECOND) OF CONTRACTS § 154. Sometimes this will place a higher burden on the promisor than morality might.

I put these examples of divergence aside. The divergence they exhibit is less stark than the examples I consider in the body of this Article. Contracting promisors often bear the burden of interpretative ambiguity, but not always. The promisee bears the burden when there is mutual misunderstanding, causing the contract to be dissolved. *See id.* § 152. Moreover, promisors will be bailed out in those cases of impracticability that threaten the survival of their operations. *See id.* § 261. These forms of divergence, unlike the ones on which I focus in the body of the Article, seem to be the product of the distinctive function of a large legal system to provide clear, predictable rules for conflict resolution and definitive adjudication when informal, more contextual mechanisms fail. Their content is explained by the effort to place the burdens of such stark rules, when necessary, on the parties who are more able to protect themselves ex ante against those burdens. The divergences I discuss in the body of the Article do not share these features.

[18] I discuss the central significance of promising to the moral agent in Shiffrin, *supra* note 7. My own approach does not take promises and their main moral force as resting on or deriving from a social convention, though many aspects of how we signal, understand, and fill in promissory gaps are, of course, based on local conventions. Most of this Article's points, however, do not depend on the rejection of conventionalism but only on the moral rules of promising being understood in the broad terms I describe.

[19] *See also supra* notes 7–9 and accompanying text.

ments that agents may form — some of which are promises, some of which are not, some of which have many but not all of the features of promises, and some whose nature is unclear. There are live philosophical and linguistic puzzles about promising, as well as gray areas and cases along a spectrum, just as with many other moral activities. Often, subtle, contextual distinctions must be deployed within the practice and along its boundaries — to do such things as differentiate promises from mere declarations of intention in context given that we do not always require invocation of the phrase "I promise" to make a promise. This Article works with clear cases of promises and puts aside these genuine unclarities, however, because they do not bear on the issues with which the Article is concerned. For the same reason, this Article also brackets much of the complexity, nuance, and qualification within contract law in favor of a simpler, perhaps overly blunt, treatment of its central doctrines.

Finally, the main argument does not presuppose a particular theory about the purposes of contract law. Hence, it begins without a detailed account of contract law's purposes. Rather, it starts from contract law's explicit self-representation of its relationship to promising to explore how well this self-representation sits alongside the moral agent's commitments.

U.S. contract law represents that a contract is an enforceable promise. Contracts do not merely resemble promises in that both involve voluntary agreements, usually concerning future activity, and often use identical language. (Later, I return to the significance of the close resemblance itself.) In U.S. law, promises are embedded within contracts and form their basis. The *Restatement of Contracts* defines a contract as "a *promise* . . . for the breach of which the law gives a remedy, or the performance of which the law in some way recognizes as a duty."[20] The *Restatement*'s definition of a promise is not technical. It invokes the familiar notion of the communication of an intention, the content and context of which justify the recipient in believing that a commitment has been made through its communication.[21] The language of promises, promisees, and promisors saturates contract law — in decisions, statutes, and the *Restatement*. It also permeates the academic literature through its common characterization of contracts as the law of enforceable promises and by its formulation of the foundational questions of contract as which promises to enforce, why, and how.[22] Notably, in U.S. law, promises of the right sort may form the

[20] RESTATEMENT (SECOND) OF CONTRACTS § 1 (emphasis added).

[21] *See id.*

[22] *See, e.g.*, P.S. ATIYAH, PROMISES, MORALS, AND LAW (1981); FRIED, *supra* note 2, at 7–14; ROSCOE POUND, AN INTRODUCTION TO THE PHILOSOPHY OF LAW 133–68 (rev. ed. 1954); W. DAVID SLAWSON, BINDING PROMISES 173 (1996); Morris R. Cohen, *The Basis of*

basis for a contract without an additional intention to enter into a legally binding arrangement.[23] Suppose we start by taking the law's self-description seriously and conceive of contracts as resting upon promises per se. As I argue below, a virtuous agent could not accept this self-description as accurate while also accepting the justification and structure of some of the divergence of contract from morality.

A. The Divergence Between Promise and Contract

As I have already observed, U.S. contract law diverges from the morality of promises. Contract law would run parallel to morality if contract law rendered the same assessments of permissibility and impermissibility as the moral perspective, except that it would replace moral permissibility with legal permissibility[24] and it would use its distinctive tools and techniques to express and reflect those judgments. For example, typically, a promisor is morally expected to keep her promise through performance.[25] Absent the consent of the promisee, the moral requirement would not be satisfied if the promisor merely supplied the financial equivalent of what was promised. Financial substitutes might be appropriate if, for good reason, what was promised became impossible, or very difficult, to perform. Otherwise, intentional, and often even negligent, failure to perform appropriately elicits moral disapprobation. If contract law ran parallel to morality, then contract law would — as the norms of promises do — require that promisors keep their promises as opposed merely to paying off their promisees. The only difference is that it would require this as a legal, and not merely a moral, matter.

1. Specific Performance and Damages. — Contract law, however, diverges from morality in this respect. Contract law's dominant remedy is not specific performance but expectation damages. Usually, the financial value of the performance is demanded from the promisor, but actual performance is not required (even when it is possible), except in special circumstances.[26] Further, intentional promissory breach is not subject to punitive damages,[27] that is, to those legal damages that express the judgment that the behavior represents a wrong.[28] Notably, U.S. law typically makes damages for emotional distress and attorney's fees unavailable upon breach.[29]

There are two further examples of the divergence over the significance of performance. First, one cannot obtain an order of specific performance even when one successfully alleges anticipatory repudiation. Even prior to the directed time of performance, a court is unlikely to direct specifically that the promised performance should occur. On the contrary, moral observers would direct that the performance should occur as promised, unless the promisee waives. The difference between the moral and legal reaction to breach does not appear only after the specified time for performance elapses.

Contract, 46 HARV. L. REV. 553, 571–92 (1933); Craswell, *Two Economic Theories*, *supra* note 3, at 19–44; Melvin Aron Eisenberg, *Donative Promises*, 47 U. CHI. L. REV. 1, 1 (1979); Melvin A. Eisenberg, *The Theory of Contracts*, *in* THE THEORY OF CONTRACT LAW, *supra* note 3, at 206, 240–64; Charles J. Goetz & Robert E. Scott, *Enforcing Promises: An Examination of the Basis of Contract*, 89 YALE L.J. 1261 (1980); James Gordley, *Contract Law in the Aristotelian Tradition*, *in* THE THEORY OF CONTRACT LAW, *supra* note 3, at 265, 268–97.

[23] *See* RESTATEMENT (SECOND) OF CONTRACTS § 21 ("Neither real nor apparent intention that a promise be legally binding is essential to the formation of a contract, but a manifestation of intention that a promise shall not affect legal relations may prevent the formation of a contract."). In other countries, the presumption sometimes runs the other way. For instance, in British law, the intention to enter into legal relations is presumed for promises between commercial entities but must be positively proved for social and familial promises to become contracts. *See* THE ENFORCEABILITY OF PROMISES IN EUROPEAN CONTRACT LAW 113 (James Gordley ed., 2001). American law also evinces reluctance to enforce certain sorts of familial promises, though this resistance is not articulated through a doctrine that there must be additional intent to enter into legal relations.

[24] Legal impermissibility would substitute for moral impermissibility, legal requirement for moral requirement, and so on.

[25] *See* FRIED, *supra* note 2, at 17.

[26] *See, e.g.*, U.C.C. § 2-716 (2005) (explaining that specific performance may be available when goods are unique, the buyer cannot find cover, and under other "proper circumstances"); RESTATEMENT (SECOND) OF CONTRACTS § 359(1) (explaining that specific performance is not available when damages adequately protect the expectation interest); 5A ARTHUR L. CORBIN, CORBIN ON CONTRACTS § 1139 (1964) (discussing specific performance as only a supplemental remedy).

[27] *See, e.g.*, RESTATEMENT (SECOND) OF CONTRACTS § 355. Punitive damages are available in cases of fraud or other violations of contract law that are also torts. For a brief period, California awarded punitive damages for "bad faith" contractual breach. But this doctrine covered only the very narrow class of cases in which the breaching party not only breached, but also contested liability without a good faith belief in the defense. Seaman's Direct Buying Serv., Inc. v. Standard Oil Co., 686 P.2d 1158, 1167 (Cal. 1984); *see also* Nicholson v. United Pac. Ins. Co., 710 P.2d 1342, 1348 (Mont. 1985). *Freeman & Mills, Inc. v. Belcher Oil Co.*, 900 P.2d 669 (Cal. 1995), sharply limited the doctrine to the insurance context a decade later. Even under *Seaman's*, intentional breach with a ready payment of compensatory damages would still have been perfectly acceptable. Nonetheless, the quite modest application of punitive damages in *Seaman's* was still met with consternation. It was severely constricted after three of its authors, Rose Bird, Joseph Grodin, and Cruz Reynoso, were all unseated from the California Supreme Court in the 1986 election. Although the retention election focused on their positions on the death penalty, Professor David Slawson obliquely suggests some connection between the election and those who blamed the court's decision for higher liability insurance costs. *See* SLAWSON, *supra* note 22, at 110–11. Some states still recognize some form of bad faith breach and subject it to punitive damages. *See id.* at 112–32.

[28] *See, e.g.*, Robert D. Cooter, *Punitive Damages, Social Norms, and Economic Analysis*, 60 LAW & CONTEMP. PROBS. 73, 73–74 (1997) (stating that punitive damages represent legal recognition of a serious wrong).

[29] *See* RESTATEMENT (SECOND) OF CONTRACTS § 353 (allowing damages for emotional distress only when breach causes bodily harm or when serious emotional distress is "a particularly likely result" of breach); 24 RICHARD A. LORD, WILLISTON ON CONTRACTS § 66:67 (4th ed. 2002) (reporting that attorney's fees are generally not recoverable in contracts cases because they are viewed as penal, that the Uniform Commercial Code does not provide for punitive damages, and that the common law rule regarding attorney's fees was not purposefully abrogated by the Uniform Commercial Code provision for consequential damages).

Second, under the *Hadley*[30] rule, promisors are liable only for those consequential damages that could reasonably have been foreseen at the time of the contract's formation.[31] From a moral perspective, this is quite strange. If one is bound to perform but without excuse voluntarily elects to breach one's duty, a case could be made that the promisor should be liable for all consequential damages. If foreseeability should limit this liability at all, what would matter morally is what was foreseeable at the time of breach rather than at the time of formation. Whereas the former reflects the idea that breach is a wrong for which the promisor must take responsibility, the latter fits better with the idea that the contract merely sets a price for potential promissory breach.

The law thereby fails to use its distinctive powers and modes of expression to mark the judgment that breach is impermissible as opposed to merely subject to a price. For this reason, I find unpersuasive the possible rejoinder that contract and promise deliver the same primary judgments — namely, that breach of promise is wrong — but that they diverge only with respect to legal and moral remedies. There are standard legal remedies (as well as legal terms) that signify that a wrong has been done. In other areas of private law, remedies such as punitive damages and specific performance are more commonly invoked.[32] Contract has a distinctive remedial regime that not only diverges from its moral counterpart, but also reflects an underlying view that promissory breach is not a wrong, or at least not a serious one.

2. Mitigation. — The mitigation doctrine provides another example of divergence. Contract law requires the promisee to mitigate her damages. It fails to supply relief for those damages she could have avoided through self-help, including seeking another buyer or seller, advertising for a substitute, or finding a replacement. As a general rule, morality does not impose such requirements on disappointed promisees. True, morality does not look sympathetically upon promisees who stay idle while easily avoidable damages accumulate. But this is a far cry from what contract expects of the promisee and what it fails to demand of the breaching promisor. Following the norms of promising, promisors would not readily expect the promisee to accept

[30] Hadley v. Baxendale, 156 Eng. Rep. 145 (Ex. 1854).

[31] *See id.*; RESTATEMENT (SECOND) OF CONTRACTS § 351(1).

[32] Despite the official distaste for punitive damages in contract, Professor Marc Galanter reports a greater rate of punitive damage awards for successful plaintiffs in contracts cases than in torts cases. Marc Galanter, *Contract in Court; or Almost Everything You May or May Not Want To Know About Contract Litigation*, 2001 WIS. L. REV. 577, 604–06. It is difficult to know what to make of his data without further details of the cases and the nature of the successful claims. As his data show, most of the punitive damage awards are provided in employment cases or in cases in which the behavior at issue was also tortious, though it was classified as a contracts cause of action. *Id.* at 605 tbl.7.

a substitute for the promised performance, at least not without a strong excuse or justification for nonperformance. Were a substitute unavoidable or justified, promissory norms would ordinarily place the burden on the promisor, rather than the promisee, to locate and provide it. It may sometimes be permissible for the promisor to *ask* the promisee to shoulder this burden when the substitute is *much* easier for the promisee to obtain or when the promisor is ill-suited to select a replacement (as when the promisee's judgment is necessary for the replacement to serve the promisee adequately). Still, even in such cases, it would usually be unacceptable for the promisor to insist were the promisee to refuse.

The difficulties in measuring and fully compensating for the costs incurred in mitigation provide another rationale for the moral stance and trouble some scholars have with respect to the current legal rule.[33] But concerns about whether compensation can be full and adequate do not exhaust the moral reasons for declining to impose a strong responsibility to mitigate on the promisee. It is morally distasteful to expect the promisee to do work that could be done by the promisor when the occasion for the work is the promisor's own wrongdoing. That expectation is especially distasteful when its rationale is that it makes the promisor's wrongdoing easier, simpler, more convenient, or less costly.

Might it be objected that the promisee, while within her rights to refuse, should not, morally, refuse a promisor's request that she mitigate? It might be thought to be stingy and overly punitive to refuse such a request.[34] If so, it might be maintained that the mitigation doctrine does in fact run parallel to morality.

Sometimes it can be morally wrong for the promisee to refuse to mitigate, especially when the costs of refusal are very steep and disproportionate to the seriousness of what is promised. But whether it is morally wrong for the promisee to refuse may depend on a number of factors to which the law is insensitive, including the closeness of the relationship, the history of the relationship, the reason for breach, the reason the promisor wants to shift the burden, and how cumbersome mitigation activities would be.

It might be suggested that the law's insensitivity to these factors is the byproduct of the need to formulate a clear rule. This is not an entirely satisfying diagnosis. The law is capable of fashioning clear, but more sensitive, rules in other equally complex contexts. Furthermore, it is unclear why, if a blunt rule is necessary, it should be fashioned to

[33] *See* Richard Craswell & Alan Schwartz, *Notes on Mitigation and Reliance by the Promisee*, *in* FOUNDATIONS OF CONTRACT LAW 64, 64–67 (Richard Craswell & Alan Schwartz eds., 1994).

[34] *See* FRIED, *supra* note 2, at 131; Melvin A. Eisenberg, *The Duty To Rescue in Contract Law*, 71 FORDHAM L. REV. 647, 654–55 (2002).

favor systematically the breaching promisor and not the promisee. Not only is the promisor the party responsible for the breach, but the wrong committed by the unreasonably reluctant mitigator-promisee is not the sort that is typically the appropriate object of legal enforcement. This wrong may fall within the category of wrongs the law should allow because interference in this particular domain might preclude recognizable realization of the virtuous thing to do — namely, to be gracious and forgiving in the face of another's wrong.[35]

3. Punitive and Liquidated Damages. — Not only are punitive damages unavailable as a response to garden-variety, intentional breach, but willing parties are not permitted to elect them in advance through legally enforceable agreements.[36] It is a delicate question whether this bar exhibits true divergence. On the one hand, agents typically cannot specify the moral seriousness of their conduct and, in particular, their misconduct: the moral status of conduct that is truly misconduct is usually independent of agents' attitudes or will. This feature of morality might lend support to the view that the rule in contracts runs parallel to morality.

On the other hand, promises occupy an interesting part of moral territory because, through them, agents themselves can alter the moral valence of some future conduct. A promise may render an action mandatory and important, when it otherwise would have been optional and, perhaps, unimportant. This created status can then be undone yet again through the consent and waiver of the promisee. Further, although I have been speaking of promises in a rather univocal way, a number of different sorts of commitments are available to agents. Parties can have tacit understandings that have many of the features of formal, strict promises. Certain sorts of affirmations or agreements constitute commitments that can be promise-like without using the most formal terms of promising. Within our moral practices of promising, agents can signify an understanding that there is a commitment but that it is fairly loose and flexible; it is not illusory, but it is subject to change for lesser reasons than would normally be acceptable for standard promises. Consider the following commitment: "I promise to be there if I can, but life is complicated right now and I can't commit for sure." The issuer is surely bound to appear if her schedule is free and her car and legs function; she is bound to turn down new, unanticipated, and conflicting requests for commitment or attendance; but she is not duty-bound to attend if it turns out that working late is necessary to meet a preexisting deadline. We are also able to signal

[35] Interestingly, French law does not impose a similar duty of mitigation on the promisee, and some acts of mitigation in the United States would be seen as breach by the promisee in France. *See* de Moor, *supra* note 3, at 106–07.

[36] *See* U.C.C. § 2-718(1) (2005); RESTATEMENT (SECOND) OF CONTRACTS § 356(1).

when such looseness and flexibility is out of order, such as when one makes a solemn commitment to be there *no matter what*. One might regard the ability to specify punitive damages as a very rough legal counterpart to the poorly defined mechanisms through which parties mark a particular promissory relation as especially serious or not. If so, then the law does show divergence by disallowing enforceable specifications of liquidated damages that exceed rough approximations of market value.[37] As I say, though, this would be a rather rough method of capturing this aspect of promising — both because it might better be captured through more clearly specified content within the contract, for example through conditions of performance, and because it marks a departure from the more general inability of moral agents to specify for themselves the significance of their own moral failures and the appropriate remedies.

B. An Objection to the Claim of Divergence

Given the plasticity of our promising practices, might the claims about divergence be challenged by arguing that contract law does not treat promises differently than does the moral system, but rather that it produces different promises with particular contents? One might attempt to recharacterize the divergent contract rules that I identify instead as rules that inform the content of what is promised between contractors. Justice Holmes famously declared that a contract to perform should be understood not as a promise to perform full stop, but as a promise either to perform or to pay damages: "The duty to keep a contract at common law means a prediction that you must pay damages if you do not keep it, — and nothing else."[38] Moreover, we might generally regard the promises that contractors enter into as implicitly incorporating the background contract law and, in turn, producing complementary terms. On this rendering, contract does not diverge from the rules of promising, but rather provides a complex background structure for certain promises, a structure that then infuses their content.[39]

[37] *See* U.C.C. § 2-718(1); RESTATEMENT (SECOND) OF CONTRACTS § 356(1).

[38] Oliver Wendell Holmes, Jr., *The Path of the Law*, 10 HARV. L. REV. 457, 462 (1897); *see also* OLIVER WENDELL HOLMES, JR., THE COMMON LAW 235–36 (Transaction Publishers 2005) (1881) (discussing the exceptionality of specific performance and noting that "[t]he only universal consequence of a legally binding promise is, that the law makes the promisor pay damages if the promised event does not come to pass").

[39] *See, e.g.*, KAPLOW & SHAVELL, *supra* note 2, at 191–92. Richard Craswell may also be read as taking this position. *See* Craswell, *Contract Law*, *supra* note 3, at 490 ("[T]he fidelity principle is consistent with any set of background rules because those rules merely fill out the details of what it is a person has to remain faithful to, or what a person's prior commitment is deemed to be.").

This rejoinder is unpersuasive for several reasons. The first, which I mention only to put aside, is that it does not pertain to the claim that the doctrine of consideration diverges from morality. This strategy can only work, if it does at all, to recharacterize gap-filling, the specific conceptions of the duties in a promise, and the responses to breaches of binding promises as specifications of the particular content of a binding promise. But it cannot address the divergence between the different *sorts* of promises that contracts and morality treat as binding, for instance their divergence over whether unilateral promises bind.

Even when the reconceptualization strategy is pertinent, it seems unpersuasive. If the promise to perform were plausibly interpreted as really a promise either to perform or to pay compensatory damages, it still would not eliminate the divergence over punitive damages. If the breaching party fails to perform *and* fails to pay damages voluntarily, there is a breach *even on the reinterpretation* of the meaning of the relevant promise. The law will respond only by providing expectation damages. It will not respond to this recharacterized promise and its breach with measures reflecting disapprobation.[40]

Should we go further and reinterpret the content of the promise as a promise only to perform *or* to pay compensatory damages, whether voluntarily or through compulsion upon legal complaint? This further recharacterization provokes another worry, namely about the assumption that the contents of promises are indefinitely plastic and utterly up to their makers. I have my doubts about this assumption. It is out of bounds to say: "I solemnly promise to do X, but I may fail to do so if something better comes along; moreover, if it does, you can only expect X's market value from me, although you may need to enlist the help of others to pry it out of my clenched fist. Further, let us now declare that should I fail, it will not be the sort of thing deserving of moral reprobation so long as eventually you are made whole monetarily. Moreover, it is not the sort of thing you may be upset with me over or view as showing my bad character." This is not a full-fledged promise. Its elaboratory remarks defy the language of its opening gambit. They clarify that it is not a promise at all, while attempting to elicit the interlocutor's acknowledgment of that fact. Rather, it seems to be the statement of an intention to act, along with an acknowledgment that the statement will, in this context, render the utterer susceptible to one sort of liability at the hands of another. But there is no commitment by the utterer to *do* anything at all. Although one can declare within a promise some of the conditions under which the promised performance may not occur, those conditions cannot coherently extend so far as to include any situation in which the promisor has a change of heart or entertains a better offer.

[40] *But cf. supra* note 27 (discussing a limited exception to the general lack of disapprobation).

Further, I doubt that one may alter by declaration or by agreement the moral significance of a broken promise (even within its terms and even if the promise itself already has a weight of its own that is partly settled by the parties). A promise may make a nonobligatory action obligatory, but only because the object of the obligation is within the promisor's power in the first place (at least for standard sorts of promises). By contrast, the power to alter the significance and appropriateness of *others'* reactions to a broken obligation is not within the power of the promisor. It does not seem to be the sort of thing that could be altered by consent or made *part* of the content of the promise. In response to another's wrong, we have the elective power to forgive, but forgiveness involves, among other things, recognition of a past wrong, not a power to make it the case that the wrong was never a wrong. Because contract's divergences involve features that are not in the power of moral agents to elect, these divergences should not be understood as components of a framework that infuses the content of certain promises.

III. IS THE DIVERGENCE PROBLEMATIC?

Although many aspects of the promissory and contractual regimes diverge from one another, the two are not flatly inconsistent. Contract law does not require promisors to breach or to respond inappropriately to breach.[41] It will not place pressure on the promisor to behave mor-

[41] Does corporate law require those with a fiduciary duty of care or of loyalty to shareholders to pursue an efficient breach if it would clearly enhance share value? The question is probably academic. The dutyholder who wishes to comply in the face of an opportunity for efficient breach could probably successfully invoke the business judgment rule by claiming that breach was not clearly in the company's or the shareholders' long-term interest, whether because of standard uncertainties in markets or reputational costs associated with intentional breach and the consequent financial repercussions, or because intentional breach conflicts with the company's mission statement. *See* STEPHEN M. BAINBRIDGE, CORPORATION LAW AND ECONOMICS 414 (2002). But suppose the dutyholder forswore these defenses or they were overcome. Could a dutyholder defend against a charge of breach of fiduciary duty for failure to pursue an efficient breach of contract on the ground that intentional breach of promise is immoral? It might depend on the jurisdiction. The general rule is that corporate directors are to maximize shareholder wealth. *See, e.g.*, Dodge v. Ford Motor Co., 170 N.W. 668, 683–84 (Mich. 1919). Many, but not all, states allow or require those with fiduciary duties to take into account third-party interests, including the interests of contracting parties and the community. *See* BAINBRIDGE, *supra*, at 414; *see also* 1 AM. LAW INST., PRINCIPLES OF CORPORATE GOVERNANCE: ANALYSIS AND RECOMMENDATIONS § 2.01(b)(2) (1992) (recommending that governance statutes permit corporations to take into account appropriate ethical considerations). The issue also arises in other contexts in which there are fiduciary duties, such as those of an executor. Professor Liam Murphy brought my attention to *Ahmed Angullia Bin Hadjee Mohamed Salleh Angullia v. Estate & Trust Agencies (1927), Ltd.*, [1938] A.C. 624 (P.C.) (appeal taken from Sing.), which held that an execu-

ally, but neither will it directly proscribe moral behavior by promisors. Nonetheless, despite their prescriptive consistency, the justifications for the divergence may violate the tenet that the virtuous agent should be able to accept the justification for the divergence.

A. Efficient Breach

A common justification for the remedies scheme in contract is that of the so-called "efficient breach." The justification may be stated in two ways. The stronger version contends that, morally, we should facilitate efficient breach because it promotes overall economic welfare and therefore overall social welfare. Take, for example, Professor David Slawson's claim: "People ought not to be liable for punitive damages merely for breaching a contract. They have done nothing wrong if they pay full compensation. Indeed, society loses if people do not breach contracts that would cost them more to perform than to pay compensation for breaching."[42] The weaker version concedes that breach of promise may indeed be morally wrong but maintains that it also can promote greater net economic gain, at least between the parties; that is to say, the economic gains to the promisor from breach may exceed the economic losses to the promisee.[43] Because facilitating efficient economic transactions is the main point of contract law and because orders of specific performance and punitive damages would deter efficient breach, contract law disallows them.[44] On this characterization, contract law is thought to have distinctive purposes that punitive damages would obstruct. Although breach may be immoral, that fact falls outside the domain of concern of contracts per se.[45]

Although it is disputed whether orders of specific performance and punitive damages are in fact economically inefficient in the specified respect,[46] for the purposes of this Article I will assume that the premise of efficient breach theory is correct. That is, I will assume that punitive damages and specific performance orders would deter some efficient breaches or would increase transaction costs in such a way that a contract law with punitive damages and more permissive specific performance rules would be less economically efficient in the specified sense than one without them.

Could the virtuous agent accept some version of efficient breach theory as a justification for these remedial rules? The stronger version of the theory would seem to be precluded by her moral commitments. A virtuous agent cannot believe both that a promise can be binding even if a better opportunity comes along that competes with fulfilling the promise *and* that breach of contract, involving breach of promise, is, all things considered, morally justified merely because it leads to (even only marginally) greater economic welfare.

---

tor never has a duty to break an enforceable contract because breach is an unlawful act. *Id.* at 635.

[42] SLAWSON, *supra* note 22, at 122 (emphasis omitted); *see also* Robert L. Birmingham, *Breach of Contract, Damage Measures, and Economic Efficiency*, 24 RUTGERS L. REV. 273, 284–85 (1969) (arguing that it would be socially desirable to encourage "[r]epudiation of obligations . . . where the promisor is able to profit from his default" after paying expectation damages); Linzer, *supra* note 2, at 115–16, 138–39 (criticizing efficient breach generally but defending, in the limited commercial domain in which people enter contracts for economic reasons, the argument that "law, economics, and arguably common sense all condone the deliberate and willful breach"); A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 939 (1998) (noting that punitive damage awards for breach of contract often result in "excessive and expensive performance . . . , thereby lowering the welfare of the contracting parties").

[43] I aim merely to describe the argument, not to endorse its notion of efficiency, which is yet another aspect of the justification that may be challenged.

[44] Some discussions frame efficiency arguments in contract in these weaker terms, although not always explicitly so. *See, e.g.*, Richard Craswell, *Contract Remedies, Renegotiation, and the Theory of Efficient Breach*, 61 S. CAL. L. REV. 629, 636–38 (1988) (discussing efficient breach in light of the goal of contract damages — to give compensation); Craswell, *Two Economic Theories*, *supra* note 3, at 27 (arguing that efficiency-based liability may not be compatible with promises, strictly construed, but that we should recognize another intermediate form of commitment that is compatible with efficiency-based liability); E. Allan Farnsworth, *Legal Remedies for Breach of Contract*, 70 COLUM. L. REV. 1145, 1216 (1970) (describing the system of contractual remedies as "heavily influenced by the economic philosophy of free enterprise" and showing "a marked solicitude for men who do not keep their promises"); Lewis A. Kornhauser, *An Introduction to the Economic Analysis of Contract Remedies*, 57 U. COLO. L. REV. 683, 686, 692 (1986) (stating that economic analyses view the purpose of contract law as the promotion of efficiency and that "the purpose of contract remedies is to induce the parties to act efficiently").

Neither characterization fits the defense of efficient breach that holds that it does not involve a breach of promise at all because breach in such circumstances is what the parties would have wanted if they had formed a complete contract. *See, e.g.*, Steven Shavell, *Contracts*, *in* THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 436, 439 (Peter Newman ed., 1998). I put this third characterization aside because it deserves a longer treatment than there is room for in this Article. In brief, the argument assumes that people would always have made the best bargain for themselves from an economic point of view had they made a complete contract and that what they did promise is determined by what rational economic actors would have promised had they focused on the relevant contingency. Both of these premises, among others, seem contestable.

[45] This description of the doctrine of efficient breach does not depend upon a comprehensive consequentialist view. For instance, an adherent might reject consequentialism as a moral position for individuals or even as the general foundation for legal doctrines, but merely posit that the purpose of the domain of contract law in particular is to create an economically efficient system for transfers and exchanges. Thus, its doctrines should be fashioned to pursue that aim, and in this limited domain consequentialist reasoning is appropriate. This more moderate description of the doctrine of efficient breach and its underlying justification thereby differs from that offered and criticized by Frank Menetrez, *supra* note 2. For a more comprehensive consequentialist view, see Craswell, *Two Economic Theories*, *supra* note 3, at 19–20, 26–34. Professor Craswell argues that economic analysis of contract has a broad scope and evaluates the effects of legal rules on a variety of parties' incentives and behaviors beyond performance, including incentives to rely, to insure, and to prepare to perform. *See id.*

[46] *See, e.g.*, Craswell, *supra* note 44, at 635; Richard Craswell, *Instrumental Theories of Compensation: A Survey*, 40 SAN DIEGO L. REV. 1135, 1145 & nn.10–11 (2003); Daniel Friedmann, *The Efficient Breach Fallacy*, 18 J. LEGAL STUD. 1, 6–7 (1989); Linzer, *supra* note 2, at 131; Ian R. Macneil, *Efficient Breach of Contract: Circles in the Sky*, 68 VA. L. REV. 947, 951–53 (1982).

The weaker version of the efficient breach theory does not, by contrast with the stronger version, hold that a party's breach of promise is morally justified, but rather that contract is indifferent to the moral status of the actions it recommends, promotes, or allows; it has distinct purposes that, on occasion, are furthered by breach of contract and so should be facilitated in those circumstances. Perhaps a moral agent should not breach promises, but this is not the concern of contract law.

Could the virtuous agent accept this argument as a justification for contract law's divergence from promise? She might have difficulty. This line of argument does not merely reject moral norms as a source of guidance for law, but harbors a strong conception of the independence of the domains — namely that a legal domain may pursue purposes that are not constrained by moral norms. It is not clear that a moral agent can accept a rule the justification of which is based on the rule's promoting the full-blown pursuit of a putatively valuable end when this full-blown pursuit would conflict with the moral agent's fundamental moral commitments.

That is, under efficient breach theory, what propels the lack of punitive damages is an affirmative normative position: agents *should* breach when it would yield net economic gain. So punitive damages must be foregone in order to make breach, and thereby a more efficient system of exchange, more likely. A virtuous agent can surely accept that there may be good aspects to wrongful breach on certain occasions. Yet, if such breach is indeed, all things considered, wrong, a virtuous agent cannot accept the economic benefits of breach as constituting a sufficient, or even a partial, contributory justification for the law's content. The challenge would be all the greater if the primary, positive justification for the law's content were the desirability of *encouraging* (and not merely making more likely) the wrongful conduct per se. In that case, the law (or its justification) would be suggesting a prescriptive recommendation to act wrongfully. It is hard to see how a virtuous agent could embrace that recommendation, whether explicit or implicit.

To be clear, the argument currently on the surgical table is not that it is none of contract law's business whether breach is immoral, so that we must therefore reject punitive damages because the only possible argument for them would be breach's immorality. That sort of argument (to which I will return) does not contain any implicit recommendation to breach. It does not, on its face, suffer the difficulty of asking a moral agent to accept the idea that it would be a good thing if people broke their promises and that we should create incentives for them to do so. How could a moral agent think both that breach of promise is, all things considered, wrong and also that it makes sense for us, as a community of moral agents, to create a system in which we attempt to encourage, however mildly, breach of promise (all the while holding out the possibility of deploying our moral condemnation of breach)?[47]

B. Distinctively Legal Normative Arguments

The root of the problem is that the efficient breach theory is driven by an underlying general *normative* position that directly conflicts with promissory norms, and not by a distinctively *legal* normative argument, by which I mean a moral argument whose range is specifically tailored to the special, normatively salient properties of law and its appropriate content and shape. An example of distinctively legal grounds of justification will illuminate the contrast between legal grounds and more general moral grounds. The reluctance to order specific performance could be justified on familiar, distinctively law-regarding grounds, such as the difficulty and expense of ordering and supervising performance by a reluctant party and, in some cases, the unseemly and disproportionately domineering nature of such state-enforced orders on individuals.[48] These are often persuasive grounds, though less so when performance merely involves the transfer or receipt of nonunique manufactured goods or the provision of services by representatives of a firm rather than by any particular individuals or employees. Whether or not these grounds are persuasive, they are distinctively legal. They do not question the general proposition that specific performance is the appropriate moral response to breach or anticipatory repudiation; rather, they resist the idea that specific performance should be implemented through legal means because of distinctive features associated with law and legal mechanisms.

---

[47] Why not condemn the efficient breach recommendation simply on the grounds that it is morally wrong? Why make the more complicated appeal to what justification the moral agent could accept? First, appeal to the moral agent involves a narrower and differently contoured sensitivity to the role of moral judgments in legal justification. Direct appeals to moral judgments might authorize eliminating certain options merely on the grounds that such options were morally bad ones. An accommodationist perspective need not accept that argument. A moral agent could not accept a rule that generated such morally bad options by direct appeal to their moral qualities but *could* accept as a reason that autonomous agents should have the opportunity to decide for themselves what to do. Second, the appeal to the moral agent reflects the correct order of explanation. The efficient breach justification is not rejected because it is morally wrong but because it cannot be endorsed by a moral agent and is therefore inconsistent with the imperative to accommodate. Arguments for accommodation do not appeal directly to the correctness of the position of moral agents but rather to the essential importance of morality to moral agents and the significance of their character traits for the flourishing of just institutions and cultures. As with religious accommodation, these grounds are compatible with greater neutrality toward the correctness of substantive moral views than is an approach that engages in more direct moral evaluation.

[48] *See, e.g.*, STEPHEN A. SMITH, CONTRACT THEORY 400–01 (2004) (discussing the liberty objection to specific performance).

Are there further distinctively legal grounds that would justify the divergences I identified earlier?

1. Liquidated Damages and Punitive Damage Agreements. — One might support the rule against punitive damage agreements on the distinctively legal grounds that they circumvent the state's monopoly on punishment. The law asserts a monopoly on punishment to prevent vigilantism and to ensure that punishment is meted out fairly and manifests horizontal equity. Legally administered punishment is also supposed to express the voice of the community. Punitive damage agreements allow parties privately to determine appropriate levels of punishment and then to commandeer the legal system to administer the punishment; this may threaten the interest in horizontal equity and in the community's authority to determine appropriate, proportionate responses to wrongs. Perhaps, then, the ban on punitive damage agreements might be justified on the grounds that contract law should not provide the means for private parties to circumvent the limits on private self-help established by tort and criminal law.

By relying on distinctively legal reasons, this theory avoids being in direct tension with commitments presupposed by promissory norms. It may not be entirely successful, however, in providing a complete explanation for the ban on punitive damage agreements. Whether this argument succeeds depends on an issue I will identify here but not attempt to resolve; namely, whether the social interest is in a monopoly on punishment, narrowly understood, or in exclusively determining the wider range of all remedial reactions to legal wrongs and breaches. If the social interest is in preventing the circumvention of the monopoly on punishment, a more limited means than a complete ban on contractual punitive damages would suffice. For instance, one could forbid those agreements that specify alternative damages, whether stronger or weaker than what the law otherwise provides, for independently tortious or criminal activity, but still allow other sorts of punitive damage agreements concerning mere intentional breach. If the social interest is in exercising authoritative and exclusive judgment over the significance of and reactions to breaches of law, however, then the wider ban on all punitive damages agreements makes more sense.[49]

As for the argument about horizontal equity, it would seem strange sometimes to require punitive damages to be administered in horizontally equitable ways[50] but then to permit great variation by the election of the parties. This argument faces obstacles, however. First, concerns about horizontal equity may have less traction when the parties agree specifically to the penalty. Second, variable punitive damages may make more sense in contract because the seriousness of the promise can vary between parties; variable punitive damages may represent a rough means by which to mark this difference, although in ways that blur the line between the substance of the promise and the remedy for breach.[51] By contrast, one might think that the significance of torts and crimes are less subject to manipulation or alteration by the parties involved. This is not a straightforward matter by any means. Consent may transform some actions that would be torts or crimes into legal activities. Still, if nonconsensual activity that comprises a tort or crime occurs, the parties cannot transform the moral seriousness of such nonconsensual, wrongful activity by agreement (whether ex post or ex ante). Agreement can play some role in establishing how serious a promise is, but it cannot, as I argued earlier, alter whether breach is a wrong.

Finally, to succeed, this defense of the ban on punitive damage agreements would have to distinguish between liquidated damages and consideration. The ban on punitive damages, put roughly, disallows liquidated damages that exceed approximated expectation damages; the doctrines of adequacy of consideration and unconscionability are far more permissive about inequitable exchanges, however, and allow consideration that may patently exceed the value of what is received in return. This creates a difficulty for any defense of the ban on punitive damage agreements, because many punitive damage agreements can be recast as forms of consideration. Graduated payment schedules that appear in the body of a contract may present alternative courses of performance and thereby appear to be complex articulations of enforceable contractual duties, even though they achieve the same result as liquidated damage clauses that overreach and are invalidated as penalties.[52] Some theory of what makes a voluntarily elected penalty a penalty would have to be provided to vindicate a defense of the ban. Perhaps that can be done.[53] If these challenges could indeed be met, the defense would be permissible from an accommodationist perspective because its appeal to distinctively legal normative considerations

[49] *See supra* pp. 728–29 for a discussion of the parallel moral phenomenon that agents cannot alter by declaration the moral significance or seriousness of immoral behavior.

[50] *See, e.g.*, State Farm Mut. Auto. Ins. Co. v. Campbell, 538 U.S. 408, 428 (2003).

[51] *See supra* section II.A.3, 726–27.

[52] *See* FARNSWORTH, *supra* note 1, at 817–18.

[53] I sketch some criteria in Seana Valentine Shiffrin, *Are Credit Card Late Fees Unconstitutional?*, 15 WM. & MARY BILL RTS. J. (forthcoming Dec. 2006), and suggest that determining whether a term is a penalty requires analysis of the form of the contract, its wording, and its primary function. For example, late charges should be understood as penalties if these charges do not represent the point of the exchange and are framed as responses to a failure to perform another duty that is the point of the contractual relationship.

makes it perfectly compatible with acceptance by a morally virtuous agent.[54]

2. Gift Promises and Consideration. — A different sort of distinctively legal argument appeals to the special hazards associated with legal rules. Professor Melvin Eisenberg and others make arguments of this sort about gift promises. Professor Eisenberg argues that legal enforceability of unilateral promises would cast doubt upon whether their performance was motivated by altruism and care, or by concern about legal liability.[55] This argument has the right structure because it specifically concerns the *legal* status of gift promises. Nonetheless, it seems only partly to justify the consideration rule.

First, many unilateral promises are not tendered as gifts, at least not in the sense presupposed by the argument. They are not always proferred as purely altruistic measures, designed in part to begin, reinforce, or symbolize a particular sort of intimate or special relationship. Second, the argument assumes that *performance* of the promise must be motivated by altruism and care in order for the value of the gift to be realized. Why doesn't the voluntary nature of the offer or promise of the gift sufficiently realize the purely voluntary component of gift promises? Why must its delivery also remain voluntary to achieve the values associated with gifts?

Professor Eisenberg's argument seems motivated, in part, by an effort to preserve the meaning of the gift for the recipient. But it is unclear that the existence of a legal enforcement mechanism would undo or cast significant doubt upon the motivations for compliance with gift promises between intimates. Often, in these circumstances, the motivations for compliance are fairly transparent, especially between parties who already share a special relationship. Further, the existence of legal remedies is unlikely to introduce muddiness. Typical transaction costs and risks make it rather unlikely that promisees will sue for breach for most sorts of gift promises. This is known to both parties, rendering it implausible that the promisor's motivation for fulfilling the promise is fear of legal enforcement activities and implausible for the promisee to worry that this is the promisor's motivation. Even putting aside standard transaction costs, legal enforcement is still unlikely because its initiation by the promisee would often do further damage to the underlying relationship — damage that may be disproportionate to the contemplated breach. These hazards are also likely to be known by both parties, again often affecting what motivations

[54] This defense would extend only to the ban on punitive damage agreements; it would not explain the general ban on punitive damages as a remedy in contract.

[55] *See* Eisenberg, *The Theory of Contracts*, *supra* note 22, at 230; Melvin Aron Eisenberg, *The World of Contract and the World of Gift*, 85 CAL. L. REV. 821, 846–52 (1997); *see also* KIMEL, *supra* note 2, at 46–49, 72–74; Gordley, *supra* note 22, at 330.

are in play and what motivations are surmised.[56] Although Professor Eisenberg's defense provides another good example of a distinctively legal argument, I remain unconvinced of its details and application.

C. Is the Divergence Objectionable?

The presupposition that divergent contract rules are suspect and in need of distinctively legal justification might be challenged in two ways. First, one might claim that the line between moral reaction and legal reaction captures exactly the appropriate level of concern about breach of promise.[57] Promissory breach merits personal disapprobation but not necessarily the community's concern. Second, one might press the view that law and morality occupy separate spheres. Perhaps there is a presumption against the law issuing prescriptives that directly contradict moral requirements, but there is no further presumption that the law should exhibit parallelism with moral norms.

A direct, comprehensive answer to these objections might take another article. But a taste of that answer might be gleaned by asking a more internal question about the relationship between different doctrinal areas — namely, what explains why tort and criminal law levy penalties but contract law does not? It cannot be sufficient to argue that tort and criminal law offer penalties in response to legal wrongs, not moral wrongs or at least not moral wrongs as such. This just raises the question why breach is not a legal wrong. That question is essentially a variation on the initial question and does not move us further along.

One might claim that tort and criminal law concentrate on a special set of cases involving especially bad behavior, often involving the infliction of physical harm. Tort and criminal law address a distinct range of moral wrongs, deserving of greater punishment than the decentralized, unofficial moral system can safely deliver. It might further be argued that the prevention and condemnation of physical harms serve special moral ends: they are ends that form some of the impetus for a legal system and that must be served for the system to function; further, they are relatively uncontroversial ends endorsed by a wide range of moral views.

[56] *Cf.* Anthony J. Bellia, Jr., *Promises, Trust, and Contract Law*, 47 AM. J. JURIS. 25, 34 (2002) (arguing that legal enforcement is not a perfect substitute for voluntary performance). Professor Bellia also observes that legally enforceable promises may help to form the basis for a relationship that then becomes dependent on other sources of trust. Further, even in solid interpersonal relationships, an offer to make a legally binding promise may help to reinforce the relationship of trust. *Id.* at 36.

[57] *See* SMITH, *supra* note 48, at 419–20 (endorsing in part and criticizing in part this suggestion).

Let me take these points in turn. Tort and criminal law do not specialize in physical harms only — think of white collar crimes, fraud, and defamation. Why are these wrongs palpably worse, necessarily, than intentional breach of trust? The answer is not obvious. Nor can it be plausibly maintained that physical security uniquely serves what is needed for social or legal functioning. Confidence that one can, by and large, trust others' word and their professed commitments is also essential to harmonized and civilized systems of social functioning. On the direct question of whether the mainstay of tort and criminal law — physical security — is more important morally than breach of trust, I do not know how to begin to evaluate the claim. Lapses in physical security can certainly cause more dramatic, immediate trauma; lapses in fidelity and confidence in others may cause more subtle forms of social and psychological corrosion.[58]

Maybe different legal reasons explain the distinctions. Perhaps threats to physical security are more tempting and so require greater deterrence, or perhaps self-enforcement is both more tempting and more dangerous if people orient themselves toward in-kind responses. Then the prospect of strong and decisive enforcement that includes state-administered punishment could be necessary to deter destructive retaliation for physical harms but not for promissory breach. With respect to the latter, unorganized social responses backstopped by a legal compensatory regime might be both safe and optimally deterrent.

This explanation differs from those just discussed because it does not point to a moral distinction between the wrongs but to a distinction between the habits and tendencies associated with responses to those wrongs, habits, and tendencies that themselves call for different sorts of legal responses. It therefore has a distinctively legal structure, but I doubt it will succeed on the merits. It depends not only on questionable empirical claims about moral practice, but also on how we conceive the point of contracts. On the empirical front, I worry that the culture of trust and promising is fragile in subtle ways that are difficult to track; it may require greater and more explicit forms of support, such as legal recognition, because threats to it are less salient than the threats to social order posed by acts of violence.[59] But I put this aside to pursue a different point about the significance of such empirical claims.

If the purpose of contract were purely to facilitate economic exchange, say by analogy to electronic banking, or to serve that and other goals such as deterring dangerous private vigilantism, the argument just rehearsed would have some force if empirically true. But if contract has a more robust normative function, the issue is harder. That is, if contract serves a positive normative purpose and not merely an instrumental and deterrent backstopping role, then empirical facts about the minimal remedies necessary to achieve instrumental and deterrent purposes would not be sufficient to establish that minimal remedies are appropriate. The moral purposes served by contract might require remedies that reflect the wrong of breach, independent of whether such remedies serve deterrent purposes or do so in a maximally efficient way.[60]

I will return to the sort of moral conception I have in mind in Part V. Before doing so, I want to address a final concern about the divergence between promise and contract to which I have adverted but have not explained. It provides an independent reason to investigate the justifications and effects of divergence.

---

[58] One might suggest that although there is a moral consensus that promises matter and that remedial reactions to breach are appropriate, the moral remedial rules are opaque. Reasonable people may differ about the seriousness of breach or how stringent the appropriate moral remedies should be. Further, the moral significance of any promise varies substantially according to context, the situation of the parties, and their mutual understandings. Given the controverted and highly contextual nature of the moral specifics, there is reason for the legal system to adopt a conservative posture within its rules of contractual enforcement. In response: It is unclear that the moral status of many torts and their appropriate remedies is clearer and less controverted than that of breach. In any case, even if a blunt, conservative rule is called for, it is not clear why these epistemic worries support our current promisor-favoring approach rather than a promisee-favoring approach. At least in cases of intentional breach, the latter seems the more conservative approach since the promisor can avoid being subject to the burdens associated with the rule by fulfilling the promise.

[59] To be sure, it is not a simple matter whether fragile cultures gain strength from the blunt and sometimes intrusive methods of the law. Concerns that legal recognition and involvement may harm organized religion more than they help make up one of the traditional strands of argument behind the Establishment Clause. *See, e.g.*, Daniel O. Conkle, *Toward a General Theory of the Establishment Clause*, 82 NW. U. L. REV. 1113, 1181–82 (1988); Steven H. Shiffrin, *The Pluralistic Foundations of the Religion Clauses*, 90 CORNELL L. REV. 9, 42–47 (2004); William W. Van Alstyne, *What Is "An Establishment of Religion"?*, 65 N.C. L. REV. 909, 914 (1987).

[60] Some moral conceptions recognize moral ends other than economic efficiency in contract. *See, e.g.*, ATIYAH, *supra* note 22, at 68–69, 138–46; FRIED, *supra* note 2; KIMEL, *supra* note 2, at 22–27, 100–07; Randy E. Barnett, *A Consent Theory of Contract*, 86 COLUM. L. REV. 269, 296–305 (1986); Peter Benson, *The Idea of a Public Basis of Justification for Contract*, 33 OSGOODE HALL L.J. 273, 314–19 (1995); Curtis Bridgeman, *Corrective Justice in Contract Law: Is There a Case for Punitive Damages?*, 56 VAND. L. REV. 237, 260–68 (2003); Gordley, *supra* note 22, at 307–08, 327–30; Jody Kraus, *Reconciling Autonomy and Efficiency in Contract Law: The Vertical Integration Strategy*, 11 PHIL. ISSUES 420 (2001) (critiquing economic efficiency as a normative grounding for contract while arguing for an autonomy-based theory whose implementation is sensitive to efficiency considerations); Daniel Markovits, *Contract and Collaboration*, 113 YALE L.J. 1417, 1419–21 (2004); T.M. Scanlon, *Promises and Contracts*, *in* THE THEORY OF CONTRACT LAW 86, 86–93 (Peter Benson ed., 2001); Liam Murphy, Theories of Contract (Sept. 2004) (unpublished manuscript, on file with the Harvard Law School Library); Peter Benson, *The Expectation and Reliance Interests in Contract Theory: A Reply to Fuller and Perdue*, 1 ISSUES IN LEGAL SCHOLARSHIP, June 2001, art. 5, at 26–31, http://www.bepress.com/cgi/viewcontent.cgi?article=1004&context=ils. Some of these conceptions advance distinctively legal and normative arguments about contract. Others do not.

IV. CULTURE AND THE MAINTENANCE OF MORAL CHARACTER

Thus far, I have claimed that there may be problems for a system of contract that invokes and is activated by promises as such, but whose rules diverge significantly from those of promise for reasons that are not distinctively legal. I have been focusing on how moral agents should regard a justification for a legal rule that celebrates the breach of a moral commitment, even though the rule merely permits it. I stand by the idea that the justification for a legal rule should be acceptable to the moral agent without compromising her virtue, but it may seem strange to place so much weight upon the content of a justification that may in fact be known to few.

Putting aside the merits of a transparency requirement, the rule and its justification may play a role in creating a wider culture in which pressure develops not to comply with the moral commitment, whether just because it is not legally required or because the legal permission spawns cultural habits that render moral compliance precious or alien. This possibility raises a further worry about a legal regime that introduces divergent norms that apply to agents simultaneously alongside moral norms — namely, whether moral individuals can participate in both cultures without running the risk that their participation will corrode the habits and expectations associated with moral practice.[61]

I will begin with an example to help elucidate what I have in mind, although I do not think it is wise to hang too much on any particular case. I do not know whether disapproval often follows a corporate officer's conscientious objection to taking advantage of an efficient breach. But I have witnessed several conversations in which one party regarded another with incredulity for thinking that she was morally bound not to break her lease against her landlord's will for convenience, suggesting that it made her a "chump," a moral fetishist for feeling bound given that the landlord could readily (though unwillingly) find a substitute renter.[62] Related exchanges occur with respect to contractors. A promisee fumes that the contractor did not come on time or, more realistically, did not come at all, despite repeated, firm promises. Her interlocutor regards her outrage as strange, observing that no more should have been expected; contractors regularly fail to

[61] Professors Kaplow and Shavell also briefly raise this possibility, but they do not pursue it because they regard it as a factor only relevant within a welfare economics analysis, one they believe would not entail a significant change in contract rules. *See* KAPLOW & SHAVELL, *supra* note 2, at 211–13.

[62] The ethical context in this situation may be complex, I grant, especially when month-to-month leases are unavailable and parties must make year-long commitments to gain access to a requisite good, thereby depriving them of needed flexibility.

show up on time when something better comes along, so it isn't a big deal — "it's business."

To be clear, I am not advancing the empirical assertion that a weakening of promissory honor is or will be the effect of the divergence of contract and promise, if only because of the almost comic difficulties in adducing persuasive evidence and examples. You may respond to my case by saying: "What do you mean? Breaking the lease is perfectly reasonable. In fact, it's what the landlord should expect." I respond: "Aha! That only shows how deep the corrosion goes. You've been infected too!" You demur and so on.[63]

To avoid such exchanges, it may be more fruitful to retreat to a more abstract level, to ponder how human moral agents nurture and maintain their habits and dispositions of moral agency. The basic concern begins with a background supposition about good behavior and forms of habituation in thought, emotion, and behavior. Namely, a great deal of morally virtuous behavior depends upon cultivating sound instincts and habits and allowing these to guide one's behavior. Morally good agents do not and cannot consciously redeliberate about all the relevant considerations bearing on a decision on every occasion. For everyday matters, agents must often depend on past deliberations that have become encoded into their general cognitive, emotional, and behavioral reactions to moral choices. Much of this deliberation and encoding is supported directly by social institutions and influenced more indirectly by the behaviors they encourage and render salient or standard.[64] This may be especially true when the law plays (or is meant to play) a leadership role in shaping social practice. If this abbreviated account is plausible, then we should be concerned about law's assigning significantly different normative valences and expectations to practices that bear strong similarity to moral practices, especially if we expect both practices to occur frequently and often alongside each other. That is, we should be concerned that the one will influence the other, making it more difficult to maintain those habits and reactions that are essential to the moral behavior. To expect otherwise, one would need to rely heavily on a clear delineation of the different behaviors and their proper contexts, as well as on our abilities to compartmentalize tightly.[65]

[63] I am not alone in worrying, however, about the decline of the culture of promising and its interrelation to legal norms and the expectations a legally shaped culture will induce. Roscoe Pound voiced similar anxieties. *See* POUND, *supra* note 22, at 159–68.

[64] *See* THOMAS NAGEL, EQUALITY AND PARTIALITY 169–79 (1991); SAMUEL SCHEFFLER, HUMAN MORALITY 133–45 (1992); Barbara Herman, *Morality and Everyday Life*, 74 PROC. & ADDRESSES OF THE AM. PHIL. ASS'N 29, 34, 36 (2000).

[65] This argument is developed and applied to the cases of compelled speech and compelled association in Seana Valentine Shiffrin, *What Is Really Wrong with Compelled Association?*, 99 NW. U. L. REV. 839, 839–55 (2005), and in Vincent Blasi & Seana V. Shiffrin, *The Story of* West

Suppose the law of legally binding agreements, through its structure or justifications, encouraged individuals to associate the conditions of binding agreements with quid pro quo exchange or to engage freely in promissory breach when breach yields only marginal economic net gains. The worry would be that these associations and behaviors would influence how the moral agent approached promises — that the divergent treatment of agreements in contract would exert a subtle influence over time on how seriously the moral agent regarded unilateral promises and how casually she regarded promissory breach. This problem may be particularly acute for those who regard the moral practice of promising as resting on a social convention, since the boundaries of that convention are not sharply defined and could well be influenced or partly constituted by the social conventions within law.[66]

Contract and promise have features that strongly trigger this general concern. Contractual agreements are entered into frequently and are a part of daily life. They bear a strong resemblance (if, *ex hypothesi*, not identity) to promises. Even if contracts are not defined in terms of promises, both contracts and promises involve voluntary agreements, can be written or oral, and may range over the same subject matter. Further, their boundaries are not especially clear, rendering it tempting to move back and forth from one set of norms to the other.

Of course, we could identify the onset of contractual relations in a clearer way to put parties on greater notice that these agreements are subject to special rules and may be treated differently than promises. Suppose that could be done and it would be worth the associated education and transaction costs to make people aware of the distinction and special rules associated with contracts. Even so, I am not sure such clarity would eliminate the difficulty, in part because, even if the distinction is underlined, the same agreement may be both promise and contract. Defining a contract as distinct from a promise does not make the contract cease to be a promise as well. Further, once the distinction is transparent, parties may explicitly ask each other for both contractual and promissory assurances. If the very same agreement is subject to both contractual and promissory norms and these norms diverge, then the difficulties reemerge. Additionally, the norms themselves, or their justifications, may point in different directions. This

Virginia State Board of Education v. Barnette: *The Pledge of Allegiance and the Freedom of Thought*, *in* CONSTITUTIONAL LAW STORIES 433, 454–75 (Michael C. Dorf ed., 2004).

[66] *See* KAPLOW & SHAVELL, *supra* note 2, at 163 (arguing that conventionalists have failed to specify the full content of the social convention and that "legal rules themselves are part of the social institution of promising"). I will not lean on this point both because I am not a conventionalist and because the problem holds for nonconventionalists as well.

may either place the agent directly in a state of conflict, or more weakly, in a position in which it is tempting to treat and regard the two as alike. If contract is more forgiving of transgressions, as I have suggested, this may exert a subtle influence to treat promises less seriously.

These factors give us some reason to be cautious about endorsing the suggestion that the problems discussed in the prior Part could be avoided merely by explicitly construing contract and promise as separate domains and by recasting contracts as entities distinct from promises. They may also give us reason to be cautious about even those divergences that can be grounded in distinctive legal normative justifications.

A. Poker and Corporate Etiquette

Of course, there are many occasions on which it is permissible to act in ways that in other contexts would be wrong. In a poker game, for instance, it is permissible to try to mislead the other players about the content of one's cards for personal gain. Not only is misleading behavior in this context permissible and consistent with the general prohibition on deception, but we do not much worry that our behavior in poker games will corrode the relevant aspects of our moral character — our resolve not to lie and to take truth-telling and candor seriously. Games provide many examples in which sharp dealing and an effort to obstruct others or to cause them to suffer loss is encouraged, whereas such conduct would be morally disallowed in other contexts. Or, to consider another context to which Professor Meir Dan-Cohen has called our attention, we do not expect sincerity from the clerk of a large corporation who thanks us for our business.[67] Why, then, shouldn't we regard the norms of contract as analogous to the norms of games or the norms in business contexts similar to Dan-Cohen's example? In these contexts, different norms of conduct govern and are not perceived to constitute a threat to our moral agency.

Games like poker and the special behaviors permitted and encouraged within them are relatively unusual activities that are fairly rigidly defined and separated from the normal course of events in life and in relationships. One may mislead only about a narrow range of topics. Within poker, these topics are limited to what cards one has and one's confidence in one's hand. Available moves in the game are not defined in terms of moral activities outside of the game. The boundaries are rigid enough that it would be inappropriate, to say the least, to ask one's fellow player, "Yes, but what cards do you really have?" Further, the game of poker and other game-like activities often require particu-

[67] MEIR DAN-COHEN, HARMFUL THOUGHTS 247–49 (2002).

lar behaviors to achieve their aims. Trying to cause another to lose (by winning) is necessary for the aims of competition to be realized. This also explains, in part, why different standards of candor are applied to lawyers and those giving testimony in adversarial settings.[68]

By contrast, contracts pervade our lives. We cannot easily opt out of them or treat them as merely an occasional leisure activity. No clear boundaries delineate the realm of activities in which contracts and contractual norms may be encountered from the realm of activities in which promises and compliance with promissory norms is expected. The lack of clarity would persist even if promise and contract were explicitly declared to be separate domains. Unlike in poker and in Professor Dan-Cohen's case of the clerk, it is not inappropriate or a clunky category mistake in the case of a contractual commitment to ask for further reassurance — to ask "Do you really mean it? . . . You're really *promising*?" Contract is not taken to be a category of behavior incompatible with promising. It is hard to see how it could be, given the prevalence of and pervasive need for contractual and promissory commitments. Finally, it is not clear, especially in light of contract law's explicit invocation of the language of promising, that the aims of contract do intrinsically rely on relaxing or abandoning moral behavior. Indeed, that is partly what is at stake in this Article.[69]

The contrast between poker and contractual promises is not stark, but rather falls on a continuum. For instance, we may have reason to be wary of the professional poker player, for whom the game is not occasional but a way of life, unless she is awfully attentive to her character and to maintaining the boundaries between the game and her other relationships. Many lawyers lose their moral bearings. But professionals are not the only ones at risk. Professor Tom Grey reported to me that a group of couples he knew used to get together in the 1970s for evenings of Diplomacy, an especially long and intense war game. By contrast with poker, it involves the forging of alliances followed by their ultimate rupture. "[C]alculated lying and backstabbing" are "crucial parts of the game play."[70] After a period of time, the group had to stop meeting because the breaches of trust involved in the game were threatening their interpersonal relationships of trust outside the game.

Nor it is clear that we should be altogether casual or sanguine about the situation of Professor Dan-Cohen's clerk. It helps tremendously when the clerk truthfully represents the position or sentiments of management or the company as a whole, the represented party is sincere, and it is understood that the clerk represents another party.[71] This already distinguishes the case from the divergence between contract and promise. In the latter, the typically disallowed behavior is not directly in the service of representing someone who is acting in a standardly moral way.

It is more disturbing if the clerk lies on behalf of management. Professor Dan-Cohen objects that someone in the corporate office may have decided that thanking customers was politic but that no person at the corporation may have any real feelings of gratitude, so that the clerk is not representing anyone's gratitude at all.[72] However, although the corporation itself may not have a mental state of gratitude or any individual employee who cares, corporations do have practices, commitments, principles, and cultures. There is a difference between a corporation that behaves in an appreciative, respectful manner toward its customers and one that treats them purely as means. Representing the former as grateful can be appropriate, even if the particular agents who design and recite the script are not themselves grateful, whereas representing the latter as grateful is deceptive. In the latter case, the fact that the clerk plays a defined role does not exempt him from fault, though the clerk may bear less fault than the script's authors higher up in management.

In both cases, it helps that the clerk's role has defined boundaries — that the clerk says these sorts of things only while at work, while being paid to represent another. The clerk's personal insincerity may be partly distinguished from the case of contract and promise. But, given the dominance of work in our lives and its spillover effects on other facets of our lives and personalities, it may be regrettable that we ask workers to display personal insincerity routinely and that we have come to expect to treat and to be treated by others with personal insin-

[68] Witnesses and lawyers need not volunteer some facts that they would otherwise be required to volunteer by conventions of cooperative conversation; lawyers may withhold elements of their strategies and advocate positions they do not personally hold. The moral permissibility of these behaviors depends in part on the well-defined boundaries of the practice as well as the underlying justifications for the practice.

[69] The majority of the aims of contract law are surely compatible with moral constraints. In Part V, I sketch one more positive and unified account of the aim of contract law that dovetails with the moral norms of promising.

[70] Wikipedia, Diplomacy, http://en.wikipedia.org/wiki/Diplomacy_(board_game) (last visited Dec. 10, 2006). "A stab can be crucial to victory, but may have negative repercussions in interpersonal relations. . . . In some circles cheating is not only allowed, but also actively encouraged. Players are allowed and expected to move pieces between turns, add extra armies . . . , listen in to private conversations, change other players' written move orders and just about anything else they can get away with. In tournament play, however, these forms of cheating are generally prohibited, leaving only the lying and backstabbing which is prevalent wherever *Diplomacy* is played." *Id.*

[71] In such cases, it may even be questioned whether this is a real case of insincerity. His "Thanks for shopping here!" may be an abbreviated form of "The company thanks you for shopping here!" *See* DAN-COHEN, *supra* note 67, at 247–48.

[72] *See id.* at 248–49.

cerity. These expectations are exactly the sort of phenomena that may provoke concern about moral drift in the culture.[73]

B. Individual Versus Corporate Agents

It may be objected that I am writing as though contracting usually takes place between individual persons — especially individuals who have preexisting bonds or who are even emotionally vulnerable to one another. But many contracts are formed between businesses or other sorts of organizations. Do I really mean to suggest that the failures of businesses to adhere to promissory norms when transacting with each other have special moral significance per se and further, that these failures may inflict damage on the external promissory culture?

My answer is, yes — at least sometimes. To some extent, I share the intuition that promissory breach between businesses is of significantly less moment, although I worry that in part my intuition is the product of an overly blunt anticorporatism. It may represent a general stance that sweeps too broadly to include small businesses as well as megacorporations and uses an indiscriminate brush not well tailored to the underlying concerns about inequality and homogeneity. There may, however, be something worthy at its root, namely a reaction to the fact that a person intrinsically matters in a way that an economic construct, even one affecting and composed of people, does not. The insult of promissory breach against a business may not sting as harshly as when it is suffered by a person. Some of my reaction also reflects attention to a different frame of reference than the one I have been discussing. That is, the more permissive intuition with respect to breach between organizations may not be a response to the general rule or practice but rather may encapsulate more particularized ethical assessments of singular cases of business-to-business breach within closely competitive contexts. It may be understandable for a party to breach when its competitors have no compunctions against doing so when it is to their advantage, when refraining from doing so would place it at a severe competitive disadvantage, and when it would be very difficult to alter the terms of interaction on a reciprocal basis through unilateral action. Even if these intuitions are appropriate to singular instances given the rules that govern the context, the overall structure of these contexts themselves may be challenged as I have been suggesting.

Further useful distinctions can be drawn between individual and organizational promisors and promisees, and in particular between

[73] *Cf.* Roy Kreitner, *Fear of Contract*, 2004 WIS. L. REV. 429, 469 (arguing that contract law should "make room" for the predispositions of managers, employees, and judges for fairness and cooperative behavior).

contracts involving individuals and those involving only experienced organizational actors.[74] It may also matter whether the contractors have repeated interactions, whether they are members of the same linguistic or geographical community, and whether the contractual formation involves communication between people or merely filling out forms on the Internet. The distinctions between different types of contractual agents and different types of contractual content may bear mightily on the relevant analysis and the overall conclusions we reach. We should therefore be wary of overly general diagnoses and conclusions.

Nonetheless, contract norms that authorize or encourage intentional breach of promise for gain among organizational actors should still give us some pause. True, an organizational actor as such cannot have disappointed feelings or expectations, at least if expectations are taken to be mental states. But for reasons I explore elsewhere,[75] I am disinclined to think that the presence of such mental states is at all essential to the binding nature of a promise (although the consequences, including the parties' potential disappointment, may sometimes bear on a promise's strength and seriousness). Furthermore, although the organizations that make commitments are not persons, persons compose them. Within business transactions, individuals often make and receive promises. An individual representing the seller company may aver to another individual representing the buyer: "Bob, I promise you, a thousand cases will be delivered tomorrow." Individuals make decisions whether to honor or breach these promises. The promises are not thereby made personal, but the involvement of individuals in the acts of commitment, receipt, and intentional breach matters. Individuals solicit one another's trust in the process of forging promissory relations between their organizations.[76] Even if these promises are not

[74] *See, e.g.*, David Charny, *Nonlegal Sanctions in Commercial Relationships*, 104 HARV. L. REV. 373, 457–60 (1990); Alan Schwartz & Robert E. Scott, *Contract Theory and the Limits of Contract Law*, 113 YALE L.J. 541, 544–46, 550, 618 (2003) (distinguishing between intercorporate transactions and transactions involving individuals and less sophisticated organizations); *see also* Kreitner, *supra* note 73, at 466–74 (agreeing that efficiency concerns should be more dominant in interbusiness transactions but questioning whether those concerns should be exclusive).

[75] *See generally* Shiffrin, *supra* note 7 (discussing the solicitation of trust as a critical component of forging promises).

[76] *See* ERIC A. POSNER, LAW AND SOCIAL NORMS 150 (2000) (observing that "what appears to be an arm's-length contract between two anonymous firms is often the result of negotiations between two friends who belong to the same social club or sit on the board of the same charitable organization"); Lisa Bernstein, *Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry*, 21 J. LEGAL STUD. 115, 119–25 (1992) (describing the relationship between trading clubs and transactions in the diamond industry) [hereinafter Bernstein, *Opting Out of the Legal System*]; Lisa Bernstein, *Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions*, 99 MICH. L. REV. 1724, 1745–54 (2001) (stressing the importance of personal relations and one-on-one encounters in commercial cotton agreements).

addressed to individuals as such and the individuals delivering them are not the formal promisors,[77] there is something troubling about a legal system that encourages persons, whether representing themselves or others, to fail to take these solicitations seriously and to take different attitudes depending upon whether they represent themselves, others, or other entities.

Consider assertion, by analogy. Although promising differs from assertion in some respects,[78] they bear a close relationship to one another.[79] At least with respect to moderately serious matters,[80] one's moral obligations not to lie or mislead do not change when one represents or addresses a business or another sort of enterprise. (Perhaps the duty of forthcomingness and the degree of wrongfulness vary, but the fact that one is speaking to the representative of an organization in itself does not alter in fundamental ways the obligation to speak truthfully.) Why should the moral norms of promising be different? Both involve the solicitation of trust. True, in a large organizational structure, the party that initiates or authorizes the breach may be a different person from the one who makes the promise, whereas the liar is often the same person who makes the representation that attempts to draw upon another's trust. But this may not matter since the party who breaches is bound, via the relation of representation, by the invitation made by the promise-giver.

Of course, one can respond that the offeror should not regard her solicitation on behalf of an organization as her own, and that the recipient of this promise should not invest her trust in the offeror as she might were the promise offered to her qua individual, in a more personal context. Perhaps. However, I do not fully grasp what motivates the "should" other than perhaps a counsel of prudence. If this prescription would not be justified as an organizing principle with respect to the sincerity of assertions, even if a more relaxed attitude toward assertion would lead to financial gain, why would it be justified with respect to promising? In any case, it is unclear whether this should be a telling point for those of us who do not believe that the moral force of a promise (or the duties following assertion) depends on whether the promisee relies or expects performance to occur.[81]

---

[77] Daniel Markovits suggests, by contrast, that these considerations may be dispositive. *See* Markovits, *supra* note 60, at 1465–68.

[78] *See* Gary Watson, *Asserting and Promising*, 117 PHIL. STUD. 57 (2004) (discussing some of the differences between promises and assertions).

[79] *See* JUDITH JARVIS THOMSON, THE REALM OF RIGHTS 294–321 (1990).

[80] Here I mean only to bracket games, jokes, and cases such as Professor Dan-Cohen's, which regard the sincerity of pleasantries by representatives. *See supra* pp. 743–44.

[81] I reject the view that the promisee must rely on, or develop an expectation of, performance for a promise to be binding. *See* Shiffrin, *supra* note 7.

More troubling is the implicit suggestion that legal norms should encourage and expect recipients to respond this charily to one another's entreaties. This, of course, returns us to territory we have already visited, involving the degree of alienation the legal system should expect and encourage from its citizens in their everyday activities, including those involving work and economic exchange. A system that leans heavily on such alienation and compartmentalization is dispiriting to defend, to put it mildly.

V. TOWARD AN ALTERNATIVE CONCEPTION

My primary aim has been to develop and advance an accommodationist approach that renders the norms of interpersonal morality relevant to the shape of law, but in a distinctive way that draws on the perspective of moral agents as subjects of law. My secondary aim has been to deploy this approach to sound some alarms about the divergence of promise and contract, particularly with respect to contract's remedial doctrines. What I have argued so far may be summarized as follows: Moral agency must be accommodated either out of respect for agents' basic, reasonable interests in leading moral lives, or because a robust culture of promissory commitment is necessary for a flourishing political society. In either case, we have a political interest in ensuring that we, as a community, do not invoke and recognize promises within our political institutions while treating them in ways inconsistent with their value, through the stance we take toward them in our rationales for various rules. To be sure, our purpose in invoking promises may not be directly to support or encourage the culture of promising as such. Indeed, we may invoke promises, in part, because such invocation is convenient. The concept of a promise operates as shorthand that is readily accessible and familiar to most citizens, even those who are not legal initiates. The use of a moral concept as shorthand is one way to make legal outcomes more accessible and to facilitate transparency. Still, if we invoke promises, directly or indirectly, we have a duty, taking something of the form of a side constraint, not to act or reason in ways that are in tension with the maintenance of a moral culture of promising.

Along the way, I have also made gestures in the direction of a more positive theory of contract that would treat the conditions of moral agency and the culture of promising in a more complementary way — a conception of contract that would incorporate sensitivity to the moral culture of promising, rather than merely regarding these concerns as a constraint on the pursuit of our other purposes, such as collective wealth enhancement. I will end with tentative remarks about a distinctively legal normative conception of contract that would sit more comfortably with our moral agency.

Promises and fidelity to them do not, of course, require law in the way systems of real property require law, or at least socially recognized boundaries.[82] So what is the purpose of a legal regime dedicated to the enforcement of some subset of promises? Suppose one did not start from a purely instrumental point of view. Would generally morally compliant and highly proficient agents who are not shy about making and keeping promises have reasons to establish a system of contract?

I believe they would. In related work, I have defended the claim that in addition to the work they may do in facilitating cooperation or the pursuit of parties' ends or projects, promises play a significant moral function in interpersonal relationships.[83] Promises and their availability provide a concrete (and I believe indispensable) way for parties to reaffirm their equal moral status and respect for each other under conditions in which possibly divergent present or future interests create vulnerability. The promissory commitment represents an effort to disable and manage some of the hazardous mechanisms and effects of power, hierarchy, and vulnerability. These reasons may be extended to illuminate the function of promises between nonintimates as well. For the purposes of this Article, I assume these claims are true.[84]

One might then understand contract as the public complement to the private promissory relationship. In creating a contract, the parties render public their efforts to manage morally their disparate interests, as well as the associated latent or emergent vulnerabilities this disparity may create or feed. Creation of a contract invites this relationship to be witnessed, recognized, and scrutinized by the public.[85] The purpose of rendering the relationship public might vary according to circumstance and content. In some cases, contracts provide assurance — going public is meant to assuage concerns that one or more parties have about the security of the arrangement. Motivations like these are familiar in both business contexts and familial contexts, including the public promises involved in marriage. But the emphasis on contract as primarily a mode of assurance, a response to the worry that things may go wrong, is exaggerated.[86] Contract law may play an important function even outside nonideal moral circumstances. Parties may well seek to create contracts for reasons that are not predominantly grounded in fear, lingering distrust of their promissory partners, or even more innocuous concerns about inadvertent breach.

For other parties, by contrast, going public may be a demonstration of feelings of strong security in the relationship and in the reliability of the commitment; one or both parties may be so confident of the commitment that they are happy to render it public and regard their willingness to do so as a symbol of their good intentions. Again, motivations like these are familiar in both business contexts and familial contexts, including the public promises involved in marriage. In other cases, contract serves a positive gap-filling function; parties may come to the essence of an agreement but rely on public rules designed to provide reasonable accommodations of the parties' interests and any relevant public interest to resolve open questions.[87] In still other cases, given what is at issue in the agreement — for example, the use of important resources — going public may be mandatory because public oversight of such resources is necessary to protect broader interests.[88]

Except in cases of the latter type, why should the public attend to these commitments and expend effort to enforce them, as well as establish norms that fill in the gaps that promissory parties fail to anticipate or resolve? A partial answer refers to reasons quite familiar from our discourse about contract. First, although promises solve and manage certain dynamics of vulnerability, they also generate new vulnerabilities. The public has an interest in protecting parties from the consequences and harm caused by breaches that result from these vulnerabilities. Second, and more broadly, the reinforcement of equal status facilitated by promises takes on a political value when made public. In addition to the political interest in a culture of taking commitment seriously, there are reasons to affirm and support such public declarations of equal status and such good faith efforts to manage di-

[82] The counterclaim, insofar as it encompasses the claim that socially recognized boundaries are essential to promising, engages the debate about conventionalism and promising. *See, e.g.*, Shiffrin, *supra* note 7 (defending a nonconventionalist view of promising). *But see supra* note 66.

[83] Shiffrin, *supra* note 7.

[84] For a more complete treatment of this issue, see *id.*

[85] Why would the interest in rendering the relationship public require law? Could other forms of social disclosure perform this function? I will only gloss these important questions. In brief, some alternative forms can work, albeit in discrete, insular, and small-scale contexts. *See, e.g.*, Bernstein, *Opting Out of the Legal System*, *supra* note 76, at 115, 119–30, 132–35; *see also* Charny, *supra* note 74, at 392–97, 412, 417–19 (discussing nonlegal sanctions for breach and the limited contexts of their effectiveness). Our culture, interestingly, lacks a clear public forum other than law in which the socially cooperative community has an official voice. The need for law to serve as our collective voice may not be an essential feature of all cooperative life, although it may be an essential feature of large-scale democratic societies. In more homogeneous cultures, religious bodies may serve as an authoritative social voice, the pronouncements of which nonetheless do not have the status of (civil) law. Given our religious and other sorts of heterogeneity (as well as the economic and civil liberty structure that nurtures such heterogeneity), it may be no accident that we lack a unifying intermediate and independent institution other than law that serves as an official public forum and voice.

[86] For instance, Professor Scanlon's claim that the institution of contracts is "centrally concerned with what is to be done when contracts have not been fulfilled" and his stress on contracts as furthering the "value of assurance" reflect an overly narrow conception of the function of contract. Scanlon, *supra* note 60, at 93, 99.

[87] *See, e.g.*, Craswell, *Contract Law*, *supra* note 3.

[88] *See* Gordley, *supra* note 22, at 280.

versity and vulnerability morally. That such a system also tends to create efficient systems of economic exchange is an important side benefit that may affect many of our decisions about how to structure the institution, but only in ways complementary to our other moral purposes.

This quick articulation is admittedly vague, but it provides a flavor of a set of rationales that could supply normative, moral reasons for an institution of contract without relying upon any direct aim to enforce interpersonal morality or to encourage virtuous behavior. Contract, on this view, is not an effort to get people to act virtuously, to prompt people to keep their promises for the right reasons, to ensure that private relationships go as well as possible, or to get people to make promises when morally appropriate to do so. It is not an effort to legalize as much as possible the interpersonal moral regime of promising, but rather to provide support for the political and public values associated with promising.

Understood in this way, a variety of the divergent aspects of contract law make sense, especially those associated with evidentiary concerns. Requirements of writing — for example, the parol evidence rule or the statute of frauds — may be understood more generally in terms of the conditions of making something verifiable to outside assessors and the public. The unconscionability and public policy doctrines manifest the limits on what commitments the public can support, given the underlying purpose of supporting equality as well as our other social aims.[89] The doctrines of mistake and impracticability presuppose notions of reasonable risk that represent our sense of which endeavors and which assumptions of risk are worth our affirmation and efforts. These characterizations refer back to public, legally normative values, but they are not in implicit or explicit tension with the view that the underlying moral promises are binding.

Although this normative conception of the purposes of contract law can readily support some divergences between promise and contract, it *may* be inconsistent with some others discussed in the body of this Article,[90] such as the general unavailability of punitive damages in contract. At the least, some standard arguments for these doctrines are in tension with the maintenance of the conditions of moral agency. Given the overriding nature of our moral commitments, as well as the dependence of a well-functioning democracy on a flourishing moral culture, there may be reason to reexamine these doctrines and their justifications, and to strive for greater convergence between promise and contract.

Many legal theorists have been particularly troubled by the idea of separating criminal law, tort, or constitutional law from moral concerns. Some have been more sanguine about conceiving of contract law as an amoral domain driven by aims entirely insensitive or indifferent to the concerns of interpersonal morality.[91] I suspect that quite the opposite is true. Contract law cannot properly be regarded as an amoral domain in the least. From an accommodationist perspective, the nesting of promise into the self-conception of contract, the ubiquity of promises and contracts, and the elemental role of commitment in social life require a legal approach to contract that is deliberately sensitive to the demands of interpersonal morality.

[89] *See* Shiffrin, *supra* note 5.

[90] Of course, some legal-normative grounds may be given for the doctrine of consideration. Exchange or its promise signals to the parties (or may serve as evidence to third parties) that a promise is legally relevant. These are permissible sorts of reasons, but they do not provide convincing support for the proposition that consideration is a necessary condition of achieving such ends, especially given the risks of creating cultural confusion about the moral significance of quid pro quo requirements. *See supra* 736–37; *see also* Scanlon, *supra* note 60, at 306–07.

[91] *See, e.g.*, JULES L. COLEMAN, RISKS AND WRONGS 73–74, 192–93, 197 (1992). *See generally* Daniel A. Farber, *Economic Efficiency and the Ex Ante Perspective*, *in* THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW 54, 66–69, 79–80 (Jody S. Kraus & Steven D. Walt eds., 2000) (discussing the intellectual history of law and economics).

A CRITIQUE OF THE PROMISE MODEL OF CONTRACT

WALLACE K. LIGHTSEY*

Lawyers and most contract scholars assume that a contract is simply a promise that the law will enforce.[1] This concept of contracts as a subset of promises, a concept referred to in this Article as the "promise model" of contract, is implicit in most legal thinking about contracts, but receives an explicit, rigorous treatment in a recent work by Professor Charles Fried.[2] Professor Fried begins his book by drawing upon basic liberal notions of respect for individual autonomy to develop the principle that a promise is morally binding on the promisor. From this principle he fashions the central thesis of his book: "[S]ince a contract is first of all a promise, the contract must be kept because a promise must be kept."[3] The remainder of the book uses this thesis to explain and criticize contract doctrine.

In a book published the same year as Professor Fried's work, Professor P.S. Atiyah also compares promise philosophy and contract doctrine.[4] Like Fried, Atiyah predicates his analysis upon the promise model of contract. Professor Atiyah, however, proceeds from this predicate to develop a methodology and to reach conclusions diametrically opposed to those of Professor Fried. Often referring to legal doctrine and to the principles and values underlying that doctrine, Professor Atiyah criticizes the principal philosophical accounts of promising and decides that promises, per se, do not create moral obligations.[5] He then develops his theory

* Law clerk to Judge John Minor Wisdom, United States Court of Appeals for the Fifth Circuit, 1983-1984 term; law clerk to Chief Justice Warren E. Burger, United States Supreme Court, 1984 Term. A.B., 1979 Duke University; J.D., 1983 Harvard University.

1. *See, e.g.*, RESTATEMENT (SECOND) OF CONTRACTS § 1 (1979) [hereinafter cited as RESTATEMENT]; C. FRIED, CONTRACT AS PROMISE 17 (1981); F. POLLOCK, PRINCIPLES OF CONTRACT 1 (12th ed. 1946); 1 S. WILLISTON, A TREATISE ON THE LAW OF CONTRACTS § 1 (W. Jaeger 3d ed. 1957).

2. *See* C. FRIED, *supra* note 1.

3. *Id.* at 17.

4. *See* P. ATIYAH, PROMISES, MORALS, AND LAW (1981).

5. *See id.* at 123-29.

that a promise is an admission of a preexisting obligation on the part of the promisor. This obligation results from either the harm to the promisee or the unjust enrichment of the promisor that would occur if the promise is not kept.[6] Thus, he concludes, if there has been no reliance by the promisee and no conferral of benefit upon the promisor, there is no immorality in the promisor's revocation of his promise.[7]

Although Professors Fried and Atiyah reach opposite conclusions, both scholars derive their theories from discrepancies between contract doctrine and the promise principle (the principle that promises are morally binding). These discrepancies fall into two related general categories. The first category concerns the legal doctrine of consideration. Because of this doctrine, contract law traditionally has not considered a mere promise to be a sufficient basis for contractual obligation; instead, the promise must have been given in exchange for a valid consideration. This is true even in the bilateral executory contract, in which one promise is exchanged for another promise. In a wholly executory contract, neither reliance nor conferral of benefit has occurred. Thus the promise principle seems to be the sole reason for regarding the contract as legally enforceable. Yet even in this situation the element of exchange is a prerequisite to legal enforceability: contract law will not enforce a unilateral promise.[8] This result is irreconcilable with the promise principle. Accordingly, Professor Fried finds the doctrine of consideration "anomalous" and "internally inconsistent."[9] Conversely, Professor Atiyah sees consideration as "a profoundly moral doctrine" reflecting the judgment that a promise alone creates no moral obligation.[10]

A second and related set of discrepancies between the promise principle and contract doctrine concerns the role of the community

6. *See id.* at 184-202.

7. *See id.* at 202-15.

8. This statement is subject to the qualification that, in some circumstances, a unilateral promise that has been relied upon may be legally enforceable. *See* RESTATEMENT, *supra* note 1, § 90. In such cases of promissory estoppel, however, the basis for enforcement is not the promise, but the reliance. *See id.* comment a. *See infra* text accompanying notes 114-17 for a discussion of the question whether contract law should enforce unilateral promises that have been relied upon.

9. C. FRIED, *supra* note 1, at 35.

10. P. ATIYAH, *supra* note 4, at 3-4.

in augmenting, overriding, or modifying the obligations assumed by contracting parties. If, as Professor Fried argues, the basis of contractual obligation is the morally binding nature of promises, then contract law should accord a high degree of deference to the will of the parties as expressed in their respective promises.[11] This ideal finds but imperfect expression in contemporary contract doctrine. For example, the law often imposes certain duties without regard to the wishes of the parties—notably, the duty to mitigate damages and the duty of good faith in bargaining and in performance. In addition, courts dealing with form contracts or unconscionable contracts have been increasingly willing to disregard terms of the contract found to be against public policy or otherwise contrary to societal understandings of fairness.[12]

Other questions regarding the role of the community vis-a-vis the will of the parties arise when the terms of a contract fail to address subsequent problems of performance. These "contractual accidents" include contracts predicated on mistaken assumptions and contracts in which unanticipated difficulties have arisen. Professor Fried concedes that, in these circumstances, the promise principle provides inadequate guidance.[13] He proposes that the law deal with these situations by using a "gap-filling" approach of turning to "residuary principles of civil obligation."[14] By emphasizing the degree to which contractual planning can avoid accidents, and thus implicitly downplaying the prevalence of accidents, Fried concludes that the necessity for gap-filling does not threaten the promissory principle.[15] Conversely, Professor Atiyah emphasizes the frequency of accidents to demonstrate the paramount role of community values in shaping contractual obligations.[16]

Professors Fried and Atiyah deserve praise for their efforts. They have expended much energy disputing the morally obligatory nature of promises in attempting to clarify the implications of

11. *See* C. FRIED, *supra* note 1, at 17.

12. In a radical extension of this trend, a recent article by Professor Todd Rakoff argues that contracts of adhesion should be presumptively unenforceable. *See* Rakoff, *Contracts of Adhesion: An Essay in Reconstruction*, 96 HARV. L. REV. 1173, 1176 (1983).

13. C. FRIED, *supra* note 1, at 69.

14. *Id.*

15. *See id.* at 69-73.

16. *See* Atiyah, Book Review, 95 HARV. L. REV. 509, 516-20 (1981).

their respective positions. In the end, however, the reader interested in understanding and further developing contract theory is left with the nagging feeling that all of this energy has been expended in vain. Both scholars build their theories from the promise model of contract; both discover profound discrepancies between promise philosophy and contract doctrine. Fried uses these discrepancies to criticize contract law while Atiyah uses them to criticize promise philosophy. Neither scholar, however, explores the conclusion that seems most obvious from the existence of the discrepancies—the simple proposition that contracts are not promises.[17]

This Article examines the validity of that proposition by criticizing the promise model of contract and suggesting an alternative model. Part I of the Article discusses three related inadequacies of the promise model: (1) abstraction and oversimplification; (2) focus on the fact of obligation rather than the content of obligation; and (3) unilateralism. Because the criticisms presented in Part I are developed with a more accurate and desirable contract model in mind, Part II describes the general contours of this alternative model, which is developed around the relationship between parties to an exchange. The Article concludes by commenting on the broad implications of the differences between the two models.

I. INADEQUACIES OF THE PROMISE MODEL

A. *Abstraction and Oversimplification*

An ineluctable result of viewing contracts as promises is to abstract and oversimplify the formation and performance of contracts, and thereby to skeletonize the rights and obligations that flow between contractual partners. Promise philosophy typically focuses on a promise by *A* to *B* to do *phi*.[18] Although this simple transaction may be suitable for developing the moral implications generated by a promise, alone or abstract, it represents an atrophied model of contract—a model incapable of dealing with the issues presented by the complex relations among parties to

17. *Cf.* Raz, Book Review, 95 HARV. L. REV. 916, 921 (1982) (suggesting this conclusion).

18. *See, e.g.*, MacCormick & Raz, *Voluntary Obligations and Normative Powers*, ARISTOTELIAN SOC'Y, supp. vol. 46, at 59, 60 (1972).

contemporary contracts. In particular, abstraction and oversimplification cause the three primary inadequacies of the promise model: discreteness, discontinuity, and presentiation.

1. *Discreteness*

Modeling contract on a simple promise from one abstract individual to another emphasizes the discreteness of a contractual transaction.[19] Discreteness leads one to view the transaction separately from its surrounding social context and to ignore the identities and relations of the parties.[20] This perspective in turn produces perhaps the most serious shortcoming of the promise model: the model operates within a vacuum that induces a gross and often complete deemphasis of the nonpromissory factors that may shape contractual rights. Promise theories concentrate on the promise itself or the actual or likely effects of the promise as the fount of the promisor's duties and the promisee's rights. Actually, however, these duties and rights are created, restrained, and molded by the interplay between the contract and the surrounding social context, as well as by the development of the contractual relationship itself.

A relatively stable, developed society is a necessary precondition to the genesis of the institutions of contract[21] and promising.[22] Indeed, it is an historico-anthropological truth that contracting does not occur until division of labor and a concomitant system of exchange have been established.[23] Persons who choose to form a contract thus approach each other not as isolated, abstract individuals, but as social actors who are already linked by the general interdependence that results from specialization of labor and exchange.[24] Their expectations are profoundly influenced by the social context of their transaction. Their agreement cannot be understood meaningfully without reference to trade custom,[25] positive law,[26] and general societal norms.[27] Moreover, these external factors will often control the rights and obligations of the parties without regard to the terms of their promises.

The contractual relationship itself generates additional nonpromissory sources of rights and obligations. Prior dealings between the parties may generate strong feelings of trust or expectation that trump the terms of a party's promise.[28] Furthermore, obligations beyond the promise itself may emerge as a contractual relationship develops. As Lon Fuller illustrates:

> In a written contract an employer promises the employee at the end of the year a bonus in addition to his regular salary. This promise is accompanied by the words: 'It is expressly understood by the parties hereto that the provision for a bonus herein contained shall impose no legal liability whatsoever on the employer, and that no action at law shall be brought for its recovery, it being understood that the payment of the bonus rests entirely in the uncontrolled discretion of the employer.' There are a considerable number of cases in the United States where courts have ordered the employer to pay the bonus notwithstanding language like that just quoted. However it may be qualified by words, the expectation that the bonus will be paid enters into and conditions the parties' conduct toward one another . . . .[29]

In short, the discreteness of the promise model hides the fact

---

19. *See* I. MACNEIL, CONTRACTS: EXCHANGE TRANSACTIONS AND RELATIONS 12 (2d ed. 1978).

20. *See* L. FRIEDMAN, CONTRACT LAW IN AMERICA 20 (1965); I. MACNEIL, THE NEW SOCIAL CONTRACT 60-61 (1980); Goldberg, *Toward an Expanded Economic Theory of Contract*, 10 J. ECON. ISSUES 45, 49 (1976).

21. "The fundamental root, the base, of contract is society. Never has contract occurred without society; never will it occur without society; and never can its functioning be understood isolated from its *particular* society." I. MACNEIL, *supra* note 20, at 1-2 (emphasis in original) (footnote omitted).

22. [I]t is wrong to think that social co-operation will not occur without the institution of promising. Indeed, as we have seen, the institution of promising is a relatively late arrival in the development of modern societies. The fact is that much social co-operation can take place without promising, so long as there is a sufficient degree of trust.

P. ATIYAH, *supra* note 4, at 135. As Professor Atiyah argues, any institution of promising presupposes a societal decision that promisees are entitled to the performance of promises and that promisors are not entitled to change their minds. *See id.* at 127-28; Atiyah, *supra* note 16, at 526.

23. *See, e.g.*, E. DURKHEIM, ON THE DIVISION OF LABOR IN SOCIETY 206 (Simpson trans. 1933); *see infra* text accompanying notes 84-88.

24. *See* E. DURKHEIM, *supra* note 23, at 200.

25. *See* U.C.C. §§ 1-205(3)-(5), 2-202 (1978); *see also* E. DURKHEIM, *supra* note 23, at 215.

26. Many areas of non-contract law—antitrust law or federal anti-discrimination law, for example—may determine contractual rights and obligations.

27. *See, e.g.*, U.C.C. § 2-302 (1978).

28. *See, e.g., id.* §§ 1-205(3), 2-207(3).

29. L. FULLER, ANATOMY OF THE LAW 81 (1968).

that trust, expectations, and accompanying obligations may arise from a variety of sources other than the parties' promises, such as the general interdependence between members of a developed society, the customs and norms of the particular business or social community, similar or previous transactions, or the developing relationship between the parties.[30] One might be able to force a promissory structure upon such expectations, but this structure would be artificial and frequently fictitious. Alternatively, asserting that these sources of right and obligation are noncontractual because they are nonpromissory entails explaining contemporary contract doctrine through some "gap-filling" approach such as that propounded by Professor Fried.[31] Any approach along these lines is doomed ultimately to demonstrate the relatively insignificant role of promise in contract law.[32]

The discreteness of promises is also responsible for the "infinite regress" problem in promise philosophy. This problem concerns the question why a promisee is *entitled* to the promised performance and why his expectation of receiving the promised performance is a legitimate one. H.A. Prichard provides the classic statement of the problem: "[P]romising to do this or that action, in the ordinary sense of promising, can only exist among individuals between whom there has already been something which looks like an agreement to keep agreements . . . ."[33] If one seeks to determine how promising can get underway between abstract individuals in a state of nature, the infinite regress is an insurmountable obstacle.[34]

30. I. MACNEIL, *supra* note 20, at 74-75; *cf.* Farnsworth, *The Past of Promise: An Historical Introduction to Contract*, 69 COLUM. L. REV. 576, 604-05 (1969) (quoting W. GOLDSCHMIDT, SEBEI LAW 221 (1967):

> [Within the Sebei tribe, s]uch force as promises have is "ensured by mutual interdependence, by the operation of a market in which each party to the contract had an ultimate interest in preserving his public reputation; his very survival in the community, and certainly his social and economic advancement, depended on his fulfilling the legitimate obligations he had incurred.").

31. *See, e.g.*, C. FRIED, *supra* note 1, at 69-73, 89-90.

32. *See* Atiyah, *supra* note 16, at 516; *see infra* text accompanying notes 49-50.

33. H.A. PRICHARD, MORAL OBLIGATION 179 (1949); *see also* Robins, *The Primacy of Promising*, 85 MIND 321, 336 (1976) (expectations that a promisor will conform to the practice of promising cannot be basis for promissory obligation because "any such expectations . . . presuppose an antecedent motivation to conform to the practice").

34. *See* P. ATIYAH, *supra* note 4, at 128 ("[A] vicious circle is inescapable so long as we confine our attention to the promisor himself.").

If one's inquiry concerns contractual obligation, however, the problem is irrelevant. Once one assumes the prior existence of specialization of labor, mutual interdependence, and a system of exchange governed by the norm of reciprocity,[35] it is not difficult to envision the emergence of a practice of making contracts without any antecedent "agreement to keep agreements." Parties to a contract will entertain expectations of performance that are *prior to* and *independent of* any belief in the moral obligation to keep promises. This critical sociological fact is hopelessly obscured by the discrete view of contracts under the promise model.

2. *Discontinuity*

The abstraction and simplicity of the promise paradigm create a model of contract as a transaction with rigid and easily discernible boundaries, a transaction that "commences sharply by clear, instantaneous agreement and terminates sharply by clear, instantaneous performance."[36] This vision may be appropriate for highly discrete contracts, which closely resemble the promise paradigm. Often in modern contractual relationships, however, neither the commencement nor the performance of the contract occurs with one specific, identifiable event.[37] Regarding the formation of a contract, the distinction between a promise and a set of assurances given in contract negotiations is difficult to demarcate: when a party makes one assurance after another, exactly when has that party consented to a contractual relationship?[38] Similarly,

35. The primal role of the norm of reciprocity is discussed below. *See infra* text accompanying notes 70-77.

36. I. MACNEIL, *supra* note 20, at 15; *see* C. FRIED, *supra* note 1, at 113 ("Whether or not a person has promised is a yes or no question."); *id.* at 120-21.

37. *Cf.* Atiyah, *supra* note 16, at 520 (Fried's justification of discontinuities "falls back on the oversimple paradigm . . . and neglects cases of interpretation and difficulty").

38. The example in the text derives from Hoffman v. Red Owl Stores, Inc., 26 Wis. 2d 683, 133 N.W.2d 267 (1965). In that case, the plaintiffs sold their previous business and incurred other expenses at the request of the defendant, who repeatedly assured them that it would establish the plaintiffs as one of its franchisees. The parties eventually terminated their negotiations without completing the transaction. The Wisconsin Supreme Court affirmed a verdict for the plaintiffs on the basis of the doctrine of promissory estoppel: "A promise which the promisor should reasonably expect to induce action or forbearance of a definite and substantial character on the part of the promisee and which does induce such action or forbearance is binding if injustice can be avoided only by enforcement of the promise." *Id.* at 694, 133 N.W.2d at 273 (quoting RESTATEMENT (FIRST) OF CONTRACTS § 90 (1932)). Professor Fried agrees with the court's result, but finds its reasoning incorrect:

promises and statements of intent differ only in degree.[39] Regarding performance of a contract, only gradual differences exist between what constitutes "substantial performance" (entitling the performer to insist on his rights under the contract) and what does not.[40] These distinctions become especially blurred in an ongoing contractual relationship. Yet, under the promise model, these distinctions trigger drastically different legal consequences, a result that can hardly be defended as rational when the controlling facts differ by only a slight degree.[41] Lines must occasionally be drawn, but there should be flexibility in the lines. The promise model is incapable of generating this flexibility.

3. *Presentiation*

By deemphasizing or even disregarding nonpromissory sources of rights and obligations, the promise model requires that the future duties of the parties be determined in the present act of promising. The promise model thus seeks to bring all future contract relations into the present—to presentiate the contract.[42] Because it is impossible for contracting parties to predict and account for all future contingencies,[43] a promise can embody only a fragment of any contractual relationship.[44] As a result, the promise model is incapable of dealing satisfactorily with long-term, ongoing contractual relationships and with unanticipated circumstances.

Because it combines presentiation with discreteness, the promise model tends to address all questions of performance under the rubric of consent:[45] discreteness sifts out nonconsensual determinants of contractual responsibility, and presentiation collects all future duties into a neat package to which each party can give present assent. Consent may be an adequate basis for determining the duties of parties to a short-term, limited contract. As the length and complexity of the contractual relationship increase, however, the limited availability of information about the future and the constant state of flux that characterizes contemporary society take their toll. In such long-term relationships, "consent can play no more than some kind of a triggering role[;] equating consent to the full scope of complex planning is downright silly."[46]

Likewise, the promise model encounters difficulty in resolving issues raised by unforeseen situations, even in the context of short-term contracts. Presentiation requires all future contingencies to be anticipated by the parties' promises.[47] To allow contract law to

---

> The award of reliance damages was not a case of enforcement of a promise at all, since the parties had not reached the stage where clearly determined promises had been made. Reliance damages were awarded because Red Owl had not dealt fairly with Hoffman. It had allowed him to incur expenses based on hopes that Red Owl knew or should have known were imprudent and that Red Owl was not prepared to permit him to realize.

C. FRIED, *supra* note 1, at 24. Professor Fried neglects to explain the distinction between a promise and Red Owl's specific representations that it would establish the plaintiffs as a franchise operator of a Red Owl store in a particular location by a particular time. Any such distinction is at best elusive, at worst disingenuous.

39. P. ATIYAH, *supra* note 4, at 165-66; *see* S. WILLISTON, *supra* note 1, § 1A; *cf.* Greiner v. Greiner, 131 Kan. 760, 764, 293 P. 759, 762 (1930) ("Ritual scrupulousness is not required [to make a promise] and, generally, any manifestation, by words or conduct or both, which the promisee is justified in understanding as an expression of intention to make a promise, is sufficient.").

40. *See* C. FRIED, *supra* note 1, at 120.

41. Professor Fried's response—that there is no irrationality in this outcome because we are within the realm of deontological morality, *see id.* at 132, 154 n.1—is symptomatic of the abstraction and oversimplicity of the promise model. It is totally artificial to posit any discrete moral difference between the actions of the defendant in *Red Owl* and those of any common promise-breaker. *See supra* note 38.

42. I. MACNEIL, *supra* note 20, at 19; *see* Goldberg, *supra* note 20, at 49; *see also* C. FRIED, *supra* note 1, at 13-14 ("Your commitment puts your future performance into my hands in the present just as my commitment puts my future performance into your hands. A future exchange is transformed into a present exchange.").

43. *See* E. DURKHEIM, *supra* note 23, at 213; I. MACNEIL, *supra* note 20, at 8.

44. *See* I. MACNEIL, *supra* note 20, at 8-9; *cf.* E. DURKHEIM, *supra* note 23, at 214 ("[I]f we were linked only by the terms of our contracts, as they are agreed upon, only a precarious solidarity would result.").

45. This tendency is well illustrated by Professor Fried's treatment of mistake and frustration. Under Fried's analysis, both of these problems result from a failure of agreement between the parties. *See* C. FRIED, *supra* note 1, at 59-60. The presentiation of contractual duties and the consequently bloated role of consent in promissory analysis cause Fried to overlook a fundamental difference between the two problems: mistake concerns a factual assumption that is erroneous at the time of contract formation; frustration relates to a difficulty arising after contract formation that was not foreseen when the contract was made. This distinction implicates differences in analysis of the two problems. Generally, because a mistaken agreement is really no agreement, resolution of a mistake situation should be influenced more by notions of fairness than by the terms of the parties' contract. Conversely, because contractual relationships are aimed at planning future action, a situation of frustration may be resolved more appropriately according to the risk allocation of the parties' contract.

46. I. MACNEIL, *supra* note 20, at 49-50; *see* P. ATIYAH, *supra* note 4, at 150-51.

47. *See* I. MACNEIL, *supra* note 20, at 19, 62.

depart from the terms of the promises in resolving unforeseen situations would be to admit the influence of nonpromissory factors and thereby to undermine the basis of the promise model. Thus, unanticipated circumstances can be addressed within the promise model in two ways. One approach brings the circumstances within the promises through the use of some legal fiction—for example, presuming that the parties would have intended a certain result had they considered the situation at hand. This method, fictionalizing intent, has been discredited.[48] The other approach denies that resolving this type of problem is a matter of contract law. Professor Fried employs this method:

> The further courts are from the boundary between interpretation and interpolation, the further they are from the moral basis of the promise principle and the more palpably are they imposing an agreement.
>
> . . . .
>
> [W]hen relations between parties are not governed by the actual promises they have made, they are governed by residual general principles of law.[49]

Though logically coherent, such a "gap-filling" methodology is an unacceptable approach to contract law. If one concedes that contemporary contracts are generally long-term, complicated arrangements, one must recognize that unforeseen situations are a normal occurrence in contractual relationships. Disavowing the applicability of contract law in these situations relegates that law to a subsidiary role in contractual relationships.[50] Furthermore, the gap-filling approach ignores the possibility that, because parties have formed a contractual relationship, it may be inappropriate to apply, for example, tort rules to an admittedly unanticipated situation.[51] Professor Fried recognizes as much in discussing the role of

48. *See* C. FRIED, *supra* note 1, at 60-61.

49. *Id.* at 61, 69.

50. *See* Atiyah, *supra* note 16, at 516.

51. *See* Note, *Disengaging Sales Law from the Sale Construct: A Proposal to Extend the Scope of Article 2 of the UCC*, 96 HARV. L. REV. 470, 477 n.42 (1982) (criticizing application of strict tort liability, rather than implied warranty provisions of article 2, to lease of defective goods). General tort rules may be inappropriate or may require modification in the context of other, noncontractual relationships as well. For example, the rules concerning assault and battery cannot be applied straightforwardly to the situation of a parent's disciplining a child.

the principle of sharing:

> By engaging in a contractual relation A and B become no longer strangers to each other. They stand closer than those who are merely members of the same political community. . . . they are joined in a common enterprise, and therefore they have some obligation to share unexpected benefits and losses in the case of an accident in the course of that enterprise.[52]

The sentiment expressed in this passage is sound, but not true to the promise model.

In short, "promise" is just too simple and abstract an idea to serve meaningfully as the primary conceptual tool for analyzing contractual relationships. Thinking of contracts as promises leads one to ignore nonpromissory sources of rights and obligations, to assume that contracts begin and end with specific, identifiable events, and to force all future contractual duties into the present act of consent. The resulting view of contracts is unrealistic when compared with most modern contractual relationships. It is too rigid and simplified a view to serve as a foundation for the development of contract theory and doctrine.

B. Focus on the Fact, Rather than the Content, of Obligation

Instead of examining the substance of the obligation that results from a binding promise, promissory analysis tends to focus exclusively on the issue whether a binding promise has been made.[53] The analysis assumes that the content of the obligation simply mirrors the promise.[54] For two reasons, this aspect of promissory

52. C. FRIED, *supra* note 1, at 72.

53. For example, Professor Atiyah's conception of promise as admission relates only to the binding nature of the promise and the conclusive nature of its terms. *See* P. ATIYAH, *supra* note 4, at 178-79. The conception provides no guidance in dealing with many difficult issues that arise after the formation of a contract, such as the scope of good faith in performance.

54. *See, e.g.*, C. FRIED, *supra* note 1, at 19, 113; H.A. PRICHARD, *supra* note 33, at 169. This assumption is a result of the discreteness that is characteristic of promissory analysis:

> In determining the content of a transaction, discreteness calls for strictly limiting the sources of . . . the substantive content of the transaction, in order to sharpen the focus as much as possible. . . .
>
> Closely related to the foregoing is the need to equate the substance of the transaction with the promises, the consensual planning, creating it.

I. MACNEIL, *supra* note 20, at 61-62.

analysis makes the promise model an undesirable basis for contract theory. First, as a general matter, the significance of the fact of obligation is wholly contingent on the content of the obligation. Second, in shifting from promise (a philosophical notion) to contract (a legal creation), the promise model fallaciously assumes that particular legal rights logically imply particular legal remedies.

A binding promise signifies nothing until one knows the consequences of being bound. Even if one concedes that a simple promise by *A* to *B* to do *phi* entails an obligation to do *phi*,[55] the content of the obligation becomes increasingly unclear as one departs from this simple paradigm. With impossible promises,[56] promises of fact,[57] and promises about the future conduct of a third party,[58] the content of the promissory obligation cannot be extracted as a necessary logical implication of the terms of the promise. Moreover, the content of the promise itself is indeterminate in some instances, such as when extremely long and complicated written contracts are imperfectly understood by one or even both parties.[59] Finally, promises alone inadequately delineate contractual obligations both in nondiscrete contractual relationships and in unanticipated contractual situations.[60] In all of these circumstances, the content of obligation takes priority over the fact of obligation. The truth of this assertion is especially apparent in situations in which there is little or no content to a party's promissory obligations. For example, when a party breaches an executory contract but there has been no change in the market, the nonbreaching party has suffered no economic injury and can recover only nominal damages. In such a case, whether a binding promise has been made is simply irrelevant.

In shifting from the realm of philosophical discourse to the realm of legal doctrine, a more fundamental error arises in the notion that the content of obligation mirrors the promise. This notion is the traditional justification for the expectation measure of damages; the measure is seen as the monetary equivalent of the promised performance.[61] But contract is a legal creation and, therefore, the rights and obligations that emanate from a contract cannot be articulated without reference to the legal remedies protecting those rights and enforcing those obligations.[62] Hence it is fallacious to argue that a particular remedy flows naturally from a particular right. For example, a person who contracts that it will rain tomorrow has clearly "promised" that, if it does not rain, he will submit to whatever sanction the law imposes.[63] This is a simple case, but every contracting party intends in a very loose sense the legal consequences of a breach on his part.[64]

Furthermore, the law imposes many obligations and ignores many promises without regard to the intentions of the contracting parties. For instance, a manufacturer's disclaimer of liability for a personal injury caused by a defect in his product is ineffective in most jurisdictions.[65] Thus, a buyer's "promise" not to sue for such

55. Even this simple assertion is, of course, a subject of intense debate. *See, e.g.*, H.A. PRICHARD, *supra* note 33; MacCormick & Raz, *supra* note 18; Robins, *supra* note 33.

56. *See* P. ATIYAH, *supra* note 4, at 155-57.

57. *See id.* at 161-64; O.W. HOLMES, THE COMMON LAW 299 (1881).

58. *See* P. ATIYAH, *supra* note 4, at 164-65; O.W. HOLMES, *supra* note 57, at 299.

59. *See* P. ATIYAH, *supra* note 4, at 148-49; Atiyah, *supra* note 16, at 516-17.

60. *See supra* notes 18-52 and accompanying text.

61. *See, e.g.*, C. FRIED, *supra* note 1, at 17.

62. *See* Fuller & Perdue, *The Reliance Interest in Contract Damages* (Pt. 1), 46 YALE L.J. 52, 52-53 (1936); Atiyah, *supra* note 16, at 519. As Professor Atiyah has noted,

> American Legal Realism is generally credited with having demonstrated that concepts used by lawyers have no 'natural' meaning or delimitation. If the law uses the concept of 'promise' then what is a promise is a matter for legal definition. . . . [O]ne cannot draw from a concept, even though it is in current use, necessary legal conclusions unless one has first put the premisses into the concept.

P. ATIYAH, *supra* note 4, at 27. Similarly, whereas Professor Fried asserts that "[t]he moral force of a promise cannot depend on whether the promisee chooses to 'enforce' the promise," C. FRIED, *supra* note 1, at 41, he fails to see that the moral *content* of the promise will depend on what society considers satisfaction of the obligation.

63. P. ATIYAH, *supra* note 4, at 57 n.36; O.W. HOLMES, *supra* note 57, at 299. The positive role of remedy in defining contractual right is especially strong because of the doctrine of Hadley v. Baxendale, 9 Exch. 341, 156 Eng. Rep. 145 (1854), which limits consequential damages for breach of contract to reasonably foreseeable damages. *See* Fuller & Perdue, *supra* note 62, at 85; *cf.* G. GILMORE, THE DEATH OF CONTRACT 53 (1974) (discussing manipulability of the *Hadley* standard of foreseeability).

64. *See* C. FRIED, *supra* note 1, at 38 ("[For a promise to be binding, the] promisor must have been serious enough that subsequent legal enforcement was an aspect of what he should have contemplated at the time he promised.") (footnote omitted); *cf.* P. ATIYAH, *supra* note 4, at 151 ("[T]he total consequences of the promise are an elaborate mesh of the actual words used (particularly written words) and of the law.").

65. *See* RESTATEMENT (SECOND) OF TORTS § 402A comment m (1963-1964).

an injury is often, legally, no promise at all. Because of its simplistic notion that the content of contractual obligation mirrors the terms of the contract, the promise model focuses all attention on the fact of obligation rather than the substance of obligation. For this reason, the model is an unserviceable tool for identifying the rights and duties that a contract creates.

C. *Unilateralism*

The promise model emphasizes the individual, isolated promise as the basic unit from which contracts are constructed. Because it combines this emphasis with the unidirectional flow of obligation under a promise, the promise model produces a unilateralistic vision of contractual duty. This vision is undesirable for two reasons: it obscures reciprocity, the central norm of contractual relationships; and it induces a fragmented view of the obligations of contracting parties.

Any theory that views contracts as composed of separable promises relies upon a complimentarity, rather than a reciprocity, of rights and obligations. The distinction between complementarity and reciprocity can be summarized in the following manner:[66]

In a situation of complementarity—
1. a right (x) of *A* against *B* implies a duty (x′) of *B* to *A*; or
2. a duty (x′) of *B* to *A* implies a right (x) of *A* against *B*.

In a situation of reciprocity)
1. a right (x) of *A* against *B* implies a duty (y′) of *A* to *B*; or
2. a duty (x′) of *B* to *A* implies a right (y) of *B* against *A*.

Thus, in a scheme of complementarity "one [party's] rights are [the other party's] obligations, and *vice-versa*."[67] Complementarity is the basis of contractual obligation under the promise model because the structure of a promise is one of obligation by promisor in

66. What follows in the text is a paraphrase of Gouldner, *The Norm of Reciprocity: A Preliminary Statement*, 25 AM. SOC. REV. 161, 168-69 (1960).

67. *Id.* at 169.

favor of promisee.[68] In contrast, a relationship based upon reciprocity posits that each party has rights and obligations; receipt and delivery of performance are mutually contingent.[69]

Reliance on complementarity rather than reciprocity is another serious defect of the promise model. Reciprocity is a critically important social norm[70] underlying and enabling the development of specialization of labor and an accompanying system of exchange.[71] As the conception of contract crystallized around the relationship between parties to an exchange,[72] the norm of reciprocity continued to remain at the heart of the relationship. From that position reciprocity exercises a key role in stabilizing the relationship, for it makes the gratification of each party's needs contingent upon satisfaction of the other party's needs.[73] Additionally, reciprocity furnishes a basis upon which a contractual relationship can adjust to contingencies that could not have been foreseen at the time of contract formation.[74] Complementarity, on the other hand, exerts a destabilizing influence:

> If assumptions about egoistic dispositions are valid . . . a complementarity of rights and obligations should be exposed to a persistent strain, in which each party is somewhat more actively concerned to defend or extend his own rights than those of others. There is nothing in complementarity as such which would seem able to control egoism.[75]

The norm of reciprocity finds expression in numerous areas of

68. *See* H.A. PRICHARD, *supra* note 33, at 169.

69. Gouldner, *supra* note 66, at 169.

70. *See, e.g.*, L. FULLER, THE MORALITY OF LAW 20 (1964) (society is "held together by a pervasive bond of reciprocity"); L. HOBHOUSE, MORALS IN EVOLUTION: A STUDY IN COMPARATIVE ETHICS 12 (1906) ("[R]eciprocity . . . is the vital principle of society."); G. SIMMEL, THE SOCIOLOGY OF GEORG SIMMEL 387 (Wolff trans. & ed. 1950) (social equilibrium and cohesion could not exist without "reciprocity of service and return service").

71. Gouldner, *supra* note 66, at 169-70.

72. *See infra* text accompanying notes 84-100.

73. *See* Gouldner, *supra* note 66, at 167-68.

74. Fuller notes that parties sometimes deliberately leave contractual terms ambiguous with the expectation that those terms will become more definite as the relationship develops. "This they do because they cannot in advance foresee just what kind of reciprocal accommodation on the matters in issue will best serve their respective interests and advance their shared desire to achieve a workable frame of collaboration." L. FULLER, *supra* note 29, at 77.

75. Gouldner, *supra* note 66, at 173; *see also* I. MACNEIL, *supra* note 20, at 45.

contract law. The most obvious expression is the doctrine of consideration. According to Holmes,

> [I]t is the essence of a consideration, that, by the terms of the agreement, it is given and accepted as the motive or inducement of the promise. Conversely, the promise must be made and accepted as the conventional motive or inducement for furnishing the consideration. The root of the whole matter is the relation of *reciprocal conventional inducement*, each for the other, between consideration and promise.[76]

Perhaps because of the stabilizing role of the norm of reciprocity, the law generally has not recognized a mere promise, absent consideration, as capable of generating contractual obligations. Rather, the law has always required that the promise be part of a mutual exchange.[77] Related doctrines, such as the requirement of mutuality of obligation and the refusal to enforce a "nude pact," echo this theme. The duty to mitigate damages and other duties of cooperation also embody the norm of reciprocity. Although these two sets of doctrine—consideration and cooperation—are both central to the law of contracts, they are anomalous to the promise model. The doctrine of consideration reflects the principle that a promise alone is not enough to create contractual obligations. Duties of cooperation show that the law imposes obligations beyond the parties' promises in order to strengthen and stabilize the contractual relationship.

Another shortcoming of the promise model that results from unilateralism is the model's tendency to separate and isolate the duties of contracting parties, rather than to view the contractual relationship as a single organic unit. Professor Fried's conception of good faith in performance demonstrates this tendency: "[G]ood faith requires not loyalty to some undefined relationship but only loyalty to the promise itself—the faithful carrying out of the mutual promises that the parties, having come to understand their *separate* purposes, choose to exchange."[78] Such an attitude creates

76. O.W. HOLMES, *supra* note 57, at 293-94 (emphasis added).

77. *See* P. ATIYAH, *supra* note 4, at 2-3 (English common law); H. MAINE, *Ancient Law* 320-37 (16th ed. 1897) (Roman law); *see also* T. HOBBS, LEVIATHAN 87 (Oakeshott ed. 1957) (distinguishing contract from promise on basis that contract involves mutual exchange).

78. C. FRIED, *supra* note 1, at 88 (emphasis in original).

difficulties for the promise model when unforeseen obstacles arise to inhibit performance. Consider the highly criticized rule that unilateral modification of a contract in light of unanticipated difficulties requires "fresh" consideration.[79] Critics often lay the blame for this rule on the doctrine of consideration.[80] The chief culprit, however, may not be consideration but rather the fragmentation of contractual duties induced by the promise model. This fragmentary attitude toward contractual duties makes the modification appear as a separate promise rather than as an adjustment within the contractual relationship. A holistic view of contract as a unitary relationship would not require new consideration for such an adjustment; it would find consideration in the exchange upon which the relationship was based.[81]

To summarize Part I of this Article, it is worth emphasizing the significant degree to which the three inadequacies discussed in this Part reinforce each other. Abstraction and oversimplification screen out nonpromissory determinants of contractual rights and obligations and thus cause the promise model to focus on promise as the sole determinant of obligation. This focus causes the model not only to adopt the simplistic and fallacious notion that content of obligation mirrors the promise, but also to conceive of contract in terms of component promises. These two effects in turn lead the promise model to view contractual obligations as characterized by rigidity and discontinuity, a view which supports presentiation and further simplification of rights and duties.

II. AN ALTERNATIVE MODEL: CONTRACT AS EXCHANGE RELATIONSHIP

The promise model of contract is an abstract and simplistic con-

79. *See* Levine v. Blumenthal, 117 N.J.L. 23, 186 A. 457 (Sup. Ct. 1936), *aff'd per curiam*, 117 N.J.L. 389, 189 A. 54 (1937). The *Restatement (Second) of Contracts* repudiates this rule. *See* RESTATEMENT, *supra* note 1, § 89.

80. *See, e.g.*, C. FRIED, *supra* note 1, at 33-36.

81. Fragmentation can also be found in the doctrines summarized in §§ 73-77 of the *Restatement (Second) of Contracts*. These sections enumerate rules for determining when a particular act—e.g., performance of a legal duty—constitutes consideration. *See* RESTATEMENT, *supra* note 1, §§ 73-77. Rather than focusing on such discrete acts, contract law should inquire whether there is consideration (and therefore a reciprocity of relations) in the relationship as a whole.

struct that implies rigidly defined rights and obligations and that fails to account for reciprocal relations among contracting parties. The model cannot deal with the complex and ongoing nature of most contemporary contractual relationships and is a defective conceptual tool for understanding much of current contract doctrine. This Part describes a more accurate conception of contract—contract as the relationship that exists and develops among parties who have made a commitment to a future exchange.[82]

As used here, "exchange" does not signify discrete exchange, but exchange in the broad sense, encompassing any form of contemplated collaboration that involves a reciprocity of rights and obligations.[83] The paradigm of this model is an ongoing, developing relationship—such as the relationship between a union and a corporate management, or between two merchants engaging in continuous business dealings. This kind of relationship has boundaries that are both porous and continuous and an obligational structure that is shaped not only by the parties' wills but also by nonconsensual factors. Section A of this Part reviews the historical development of contract to demonstrate the origin of contract in exchange. Section B then outlines the general characteristics of the exchange-relationship model. Section C concludes this Part by noting a few limitations of this model.

A. *The Development of Contract*

The legal concept of contract is rooted in the exchange relationship. In Roman law the contractual transaction emerged as a variation of the sale transaction. As traced in Sir Henry Sumner Maine's *Ancient Law*, this emergence occurred in three stages.[84]

82. French law has a similar concept, "*achalandage*," which embraces "the sum of all relations created between a business man and his customers." Derenberg, *The Influence of the French Code Civil on the Modern Law of Unfair Competition*, 4 AM. J. COMP. L. 1, 4 (1955) (quoting WAELBROECK, COURS DE DROIT INDUSTRAL (1863-1867)). This concept, however, is used not to regulate the relations between the businessman and his customers, but to protect those relations from interference from other merchants. *See id.* The analogy to the exchange-relationship model of contract is therefore imperfect.

83. *See* L. FULLER, *supra* note 29, at 72; *see also* I. MACNEIL, *supra* note 20, at 86 (modern contractual relationships "involve a flow of exchanges, or often many flows at the same time, occurring in complex patterns not lending themselves to divisions into discrete periods").

84. *See* H. MAINE, *supra* note 77, at 315-37.

Early Roman law conflated contract and sale. Both were conceptualized as the same transaction; they were seen as a simple and simultaneous exchange of property for money. The second stage marked the beginning of the divergence between contract and sale. In this stage the law came to recognize the credit transaction, a conveyance of property in return for a promise to pay. As Maine notes, "Contract was long regarded as an incomplete conveyance."[85] The culmination of the law of contract occurred in the third stage, in which legal enforceability was conferred upon the bilateral executory contract.

The English common law of contract developed not from the incomplete conveyance but from the tort action of trespass on the case.[86] Despite this tort background, the basis of the contractual action was exchange: the special action of assumpsit (a variety of trespass on the case) was predicated upon misfeasance in the performance of an undertaking.[87] By the second half of the fifteenth century, assumpsit had been expanded to allow suit for nonfeasance, and at the end of the sixteenth century the action came to include bilateral executory agreements.[88]

The wholly executory contract is the closest that contract law has come to embracing the promise principle.[89] Thus, when wholly executory contracts became legally enforceable, contract appeared to be on the verge of slipping its exchange moorings: the next logical step would have been legal recognition of unitary promises. In both Roman and English law, however, consideration became a basic requirement of enforceability precisely at this time.[90]

This development was no coincidence, for exchange is the essence of the doctrine of consideration.[91] A traditional formulation of the doctrine defines consideration in terms of benefit to promisor or detriment to promisee.[92] This bifurcated definition is a bas-

85. *Id.* at 321 (emphasis omitted).

86. *See* G. GILMORE, *supra* note 63, at 140 n.228; Farnsworth, *supra* note 30, at 594-95.

87. Farnsworth, *supra* note 30, at 594.

88. *See id.* at 594-96.

89. *See supra* text accompanying notes 8-9.

90. *See* H. MAINE, *supra* note 77, at 337 (Roman law); Farnsworth, *supra* note 30, at 598 (English law).

91. *See* J. DAWSON, GIFTS AND PROMISES (1980); Farnsworth, *supra* note 30, at 598.

92. *E.g.*, S. WILLISTON, *supra* note 1, §§ 99-100, 102-104.

tardization of the core notion of consideration as exchange[93] insofar as the definition acknowledges consideration in unbargained-for reliance stemming from a non-exchange situation.[94] Similarly, use of the doctrine of consideration to exclude options and modifications of existing arrangements has been correctly identified as an inappropriate extension of the doctrine, because both options and modifications occur in the context of an exchange.[95]

Professor Fried's criticism of the doctrine turns on the inconsistency between the requirement of an exchange and the proposition that "[t]he law is not at all interested in the adequacy of the consideration."[96] But the latter proposition is not a logical outgrowth of the notion of consideration; rather, it is the accomplishment of the movement of late-nineteenth-century jurists (notably Holmes) to objectify legal doctrine.[97] It conflicts fundamentally with the exchange basis of contract and should be discarded. By making exchange the basis of contractual obligation,[98] the doctrine of consideration restricts legal enforceability to situations of reciprocity.[99] This restriction necessitates inquiry into the substantiality of the

93. *Cf.* Farnsworth, *supra* note 30, at 598 (bifurcated formulation has proved less durable than notion of exchange).

94. *See, e.g.*, Allegheny College v. National Chautauqua County Bank, 246 N.Y. 369, 373-74, 159 N.E. 173, 175 (1927) ("[T]here has grown up . . . a doctrine that a substitute for consideration or an exception to its ordinary requirements can be found in what is styled 'a promissory estoppel.' . . . [W]e have adopted the doctrine of promissory estoppel as the equivalent of consideration in connection with our law of charitable subscriptions."). The *Restatement* has wisely avoided the bifurcated definition of consideration in favor of a notion of consideration as exchange. *See* RESTATEMENT, *supra* note 1, §§ 71, 79. Promissory estoppel, which allows enforcement of a promise that has been relied upon even if the promise is not part of an exchange, is treated under the topic "Contracts Without Consideration." *See id.* §§ 82-94.

95. RESTATEMENT, *supra* note 1, § 87 comment b, § 89 comment a.

96. C. FRIED, *supra* note 1, at 29; *see id.* at 35. *But see* RESTATEMENT, *supra* note 1, § 71 comment b ("[A] mere pretense of bargain does not suffice, as where . . . the purported consideration is merely nominal."); *id.* § 79 comment c ("Ordinarily, therefore, courts do not inquire into the adequacy of consideration. . . . Gross inadequacy of consideration may be relevant to issues of capacity, fraud and the like, but the requirement of consideration is not a safeguard against imprudent and improvident contracts *except in cases where it appears that there is no bargain in fact.*") (emphasis added).

97. *See* G. GILMORE, *supra* note 63, at 41-45.

98. *See* A.W.B. SIMPSON, A HISTORY OF THE COMMON LAW OF CONTRACT 456-57, 491 (1975); *cf.* P. ATIYAH, *supra* note 4, at 193 ("[T]he promise is usually evidence that the transaction is an exchange and not a gift. . . .").

99. *See supra* text accompanying notes 76-77.

exchange.[100]

B. Characteristics of the Exchange-Relationship Model[101]

In contrast to the rigidity and simplicity of the promise model, the exchange-relationship model views contractual relationships as flexible, adaptable, and capable of incorporating rights and obligations from sources other than the promises of the parties. This alternative model not only presents a more accurate picture of modern contracts, but also provides a more powerful vehicle for the creation and sustenance of contractual relationships in contemporary society. This Section outlines the structural characteristics of the alternative model and then briefly discusses the sources of obligation under that model.

The exchange-relationship model focuses on the long-term contract as the paradigm of contractual relationships. Because of the limited availability of information about the future and the tendency of socioeconomic conditions to fluctuate, the model recognizes that a contractual relationship, to survive, must incorporate a capacity for adapting to changing conditions.[102] The model consequently posits a two-staged process for the delineation of rights

100. For example, in Williams v. Walker-Thomas Furniture Co., 198 A.2d 914 (D.C. 1964), a furniture company had structured the installment sales contract between it and a welfare mother in such a way that the company retained title to all goods purchased from it as long as there was an outstanding balance due on any one item. The company's agents also permitted Mrs. Williams to purchase an expensive stereo system, which they knew she could not afford, at a time when she was close to eliminating her outstanding balance. After Mrs. Williams defaulted on her installment payments, the company brought a replevin action to repossess all of the goods she had purchased from the company. The court of appeals condemned the company's conduct, but affirmed the trial court's judgment for the company nonetheless.

This holding is unfaithful to the doctrine of consideration and the principle of exchange upon which that doctrine rests. The court should have scrutinized the substantiality of the purported exchange. Although at a purely formal level the contract appeared to contain consideration, the relationship between seller and buyer was really one of bald exploitation rather than one of reciprocity based on exchange. The court should have voided the contract for lack of sufficient consideration.

101. As with much of the criticism presented in the first Section of Part I, discussion of the characteristics of this model of contract draws heavily from the work of Ian Macneil. *See* I. MACNEIL, *supra* note 19; I. MACNEIL, *supra* note 20; Macneil, *The Many Futures of Contracts*, 47 S. CAL. L. REV. 691 (1974). This Section is deliberately abbreviated to avoid repetition of points examined in the first Part of the Article.

102. *See* I. MACNEIL, *supra* note 20, at 50-51.

and obligations: (1) the initial, tentative definition; and (2) the continual adjustment of the relationship to unanticipated circumstances that occur as the future unfolds.[103] The model also recognizes that in many long-term contractual relationships, such as that created among stockholders by a corporate charter, individuals or entities may enter and exit the relationship while the relationship continues. This view of contract requires that the law take a flexible approach to contractual duties and, thus, that the rigidity and discontinuity of rights under the promise model be rejected.

Whereas the promise model treats the wills of the parties as the nearly exclusive source of obligation under contract, the exchange-relationship model recognizes three general sources of obligation: sources external to the relationship, the consent of the parties, and the relationship itself. Because the alternative model treats contract as a socio-legal relationship, it easily accounts for the controlling manner in which society and law shape contractual relationships. The model recognizes that general societal norms of fairness and reciprocity, positive law, and trade custom will qualify and even vitiate obligations assumed by contracting parties, and will impose obligations beyond those assumed by the parties.[104]

Although the exchange-relationship model thus accords a diminished role to consent as a determinant of rights and obligations, consent still performs a vital function in triggering obligation.[105] The alternative model sees the contractual relationship as essentially voluntary, in contrast with pure status relationships. In a pure status relationship, such as that between parent and child,[106] duties are based wholly on a position that cannot be divested vol-

103. *See id.* at 24-26; Unger, *The Critical Legal Studies Movement*, 96 HARV. L. REV. 561, 639-40 (1983). Of course, the initial definition may embrace procedures for structuring future adjustments. An example of such an arrangement is the establishment of an arbitration procedure.

104. *See generally* P. ATIYAH, *supra* note 4, at 130 ("We thus find a decline in the belief that the individual has the right to determine what obligations he is going to assume, and an increased strength in the belief that the social group has the right to impose its own solution on its members, dissent as they may.").

105. *See* I. MACNEIL, *supra* note 20, at 50; *cf.* P. ATIYAH, *supra* note 4, at 177 ("[P]romising may be reducible to a species of consent . . . .").

106. *See* E. DURKHEIM, *supra* note 23, at 207; L. FULLER, *supra* note 70, at 24.

untarily.[107] Besides triggering obligation, consent may also shape the content of that obligation.[108] This shaping function, however, is not a necessary characteristic of a contractual relationship. Because the exchange-relationship model recognizes the influence of external sources in defining contractual obligation, it must assert that no sharp distinction exists between a contract in which the parties have significant freedom to define their obligations and one in which the obligations are defined almost entirely (though *not* triggered) by nonconsensual factors. An example of the latter type of contractual relationship is marriage in nineteenth-century Anglo-American society. Consent was required to become married; yet, once married, the responsibilities of each partner were defined by law and custom.[109] Thus, consent played a vital role in triggering obligation but a trivial role in shaping it. In this respect, then, the exchange-relationship model downplays the shaping function of consent. Nevertheless, the emphasis that the model places on the capacity to respond to unforeseen circumstances dictates that consent play a significant shaping role in most contractual relationships.[110]

The relationship itself may also be a source of contractual obli-

107. *See* Gouldner, *supra* note 66, at 170; Pound, *The End of Law as Developed in Juristic Thought* (Pt. 2), 30 HARV. L. REV. 201, 211 (1917). Sir Henry Maine's thesis is that, as a society progresses, individual obligation becomes decreasingly status-based (one's status being defined largely by family and social position) and increasingly contract-based. *See* H. MAINE, *supra* note 77, at 168-70.

108. The distinction between triggering obligation and shaping obligation may underlie Joseph Raz's distinction between promises and other voluntary obligations. *See* Raz, *supra* note 17, at 930-33.

109. Similarly, Durkheim has described cultures in which individuals could freely enter into and exit from various family relationships. *See* E. DURKHEIM, *supra* note 23, at 207-10.

The recent history of the relationship between contract and marriage is an interesting one. In the 19th century, law and custom rigidly defined the obligations of husband and wife: the husband was required to support, the wife to serve. Johnson, *Contract Love*, STUDENT LAW., Feb. 1983, at 12, 14-15. Courts of this period were therefore unwilling to allow marital duties to be varied by contract. *See id.* at 16. Over the last fifty years, as the duties of married persons have become less status-oriented, courts have become increasingly receptive to contractual variation of marital obligations. *See id.* at 16-17.

110. Flexibility is further enhanced by the bifurcated focus of consent in complex contractual relationships. Rather than defining only the substance of exchange, consent is directed also toward shaping the structures and processes by which the contractual relationship will adjust to unforeseen circumstances. *See* I. MACNEIL, *supra* note 20, at 24-25, 47-50; *supra* note 103. Examples of such bifurcation may be found in corporate charters and collective bargaining agreements.

gation. A longstanding contractual relationship must be judged by standards different from those that apply to a brief or newly formed relationship. For example, in some jurisdictions, a contract between persons who have a history of close and trusted dealings may not be subject to the usual requirement that the contract be in writing.[111] As a contractual relationship develops, the collaboration gives rise to ties of community between the parties, which may generate additional obligations. Maine notes that

> [a]t the earliest dawn of . . . [Roman] jurisprudence, the term in use for a Contract was . . . *nexum*, and the parties to the contract were said to be *nexi*. . . . The notion that persons under a contractual engagement are connected together by a strong *bond* or *chain*, continued to the last to influence the Roman jurisprudence of Contract; and flowing thence it has mixed itself with modern ideas.[112]

Additionally, because contemporary contracts often involve ongoing relationships, maintenance of the relationship becomes an important constraint upon both parties[113] and thus implicates duties of loyalty and cooperation.

C. *Limitations of the Alternative Model*

Just as the promise model's focus on the discrete and abstract transaction creates weaknesses in that model, the exchange-relationship model has several limitations that arise because of its emphasis on the ongoing, nondiscrete contractual relationship. Specifically, the model has difficulty dealing with three problems of contract theory: the enforceability of wholly executory contracts,

111. *See, e.g.*, Harris v. Sentry Title Co., 715 F.2d 941, 945-48 (5th Cir. 1983) (imposing a constructive trust).

112. H. MAINE, *supra* note 77, at 314; *see also* Farnsworth, *supra* note 30, at 582 (in two African tribes, ongoing trading relationships are considered to be relationships of quasi-kinship); *cf.* United States Steelworkers of Am. v. Warrior & Gulf Navig. Co., 363 U.S. 574, 580-82 (1960) ("collective bargaining agreement" means the whole collective bargaining relation, not just the written contract); E. DURKHEIM, *supra* note 23, at 217 ("[E]xchange, as we have seen, is not all there is to a contract. There is also the proper harmony of functions concurring. They are not only in contact for the short time during which things pass from one hand to another; but more extensive relations necessarily result from them, in the course of which it is important that their solidarity be not troubled.").

113. I. MACNEIL, *supra* note 20, at 66.

rights and obligations under highly discrete contracts, and rights and obligations under unilateral promises.

1. *Executory Contracts*

The exchange-relationship model accords a significant influence to the role of nonpromissory factors in creating contractual obligations, and thus recognizes that expectations of performance may exist without any belief in the morally obligatory nature of promises. Similarly, by founding the doctrine of consideration on the norm of reciprocity rather than on the separate elements of benefit, reliance, and promise, the model need not rely on the binding nature of promises to justify enforcing executory agreements.

Nonetheless, it is not clear that the model supports the enforcement of wholly executory contracts. The norm of reciprocity obviously entails a duty not to disappoint the legitimate expectations of a contractual partner. But in a wholly executory situation, the norm itself cannot be the basis for determining whether an expectation of performance is legitimate, because the norm requires a *prior* right or duty upon which to operate. Thus, it is unclear whether this norm comes into play before there has been any reliance or conferral of benefit.

Because the exchange-relationship model focuses on ongoing relations, the characteristics of the model are not particularly useful in generating a reasoned response to the question whether purely executory contracts should be binding. The question simply does not arise in the context of an ongoing contractual relationship. In an ongoing relationship, reliance and conferral of benefit have already occurred and are occurring continuously. The best the model can do in addressing this question is to refer to actual practice as expressed through societal norms, law, and trade custom. If these sources of rights and obligations recognize wholly executory agreements as binding agreements, a contractual relationship will incorporate that proposition.

2. *Discrete Contracts*

Although one may fault the promise model for failing to deal adequately with complex and ongoing contractual relationships, one must nevertheless recognize that short-term, simple, discrete

contracts play a significant role in contemporary society. Because the exchange-relationship model broadens the field of obligational sources and relies heavily on cooperative adjustment of the contractual relationship over time, it does not provide great certainty. Indeed, a principal strength of the model is its emphasis of flexibility over certainty. The exchange-relationship model may therefore be inappropriate in many situations. For example, in a dispute over a contract for a sale of stock on a national exchange, legal inquiry into the "relationship" between buyer and seller not only would be a waste of time, but also would undermine the certainty and predictability vital to the operation of the exchange.

3. *Unilateral Promises*

The alternative model places contract in exchange and therefore cannot offer any reason to enforce a promise that is not part of an exchange. This incapacity exists even when a promisee has relied on a promise, as long as the reliance did not occur in the context of an exchange relationship.[114] This failure is difficult to justify theoretically, and is an embarrassing shortcoming in light of the fervor with which contemporary contract doctrine embraces the notion of promissory estoppel.[115]

One could perhaps argue that contract law should be concerned only with relations of reciprocity or, more specifically, with protecting the mutual trust necessary to the continuance of such relationships. A unilateral promise creates a relationship of complementary rights and obligations, whereas a relationship of reciprocity is both more stable and more adaptable than one of

114. If the promise occurs within an exchange relationship, it is not really unilateral. Many purported examples of promissory estoppel actually involve promises occurring within exchange situations. For example:

> A has been employed by B for 40 years. B promises to pay A a pension of $200 per month when A retires. A retires and forbears to work elsewhere for several years while B pays the pension. B's promise is binding.

RESTATEMENT, *supra* note 1, § 90 illustration 4. Similarly, the situation in Hoffman v. Red Owl Stores, Inc., 26 Wis. 2d 683, 133 N.W.2d 267 (1965), was that of negotiations preliminary to an exchange. *See supra* note 38.

115. "A promise which the promisor should reasonably expect to induce action or forbearance on the part of the promisee or a third person and which does induce such action or forbearance is binding if injustice can be avoided only by enforcement of the promise." RESTATEMENT, *supra* note 1, § 90(1).

complementarity.[116] But that distinction furnishes no reason for reciprocity to define the limits of contractual obligation. As Professor Fried asks,

> [W]hy is my enforceable promise to sell my brother-in-law my automobile less sterile than my promise to give it to my nephew? The law recognizes the *completed* transaction (after I actually hand over or sign over the automobile), presumably in recognition of my right to do with my property as I choose.[117]

One might also argue that promises are unlikely to occur in nonexchange situations. A person will rarely make a promise to another from whom the promisor neither has received some past benefit nor expects some future reciprocation. In such rare instances, the law can prevent injustice by turning to some other source of legal obligation. This argument seems to stretch the notion of "exchange" too far.

The only response to these criticisms may be that the characteristics of exchange relationships are sufficiently determinate and are sufficiently distinct from the characteristics of promise relationships that the two types of relationships should be dealt with by different branches of legal doctrine. Contract scholars and lawyers should be concerned with developing rules that encourage and strengthen exchange relationships. Furthermore, these rules should enable such relationships to adapt to unforeseen circumstances and to withstand the pressures of adverse conditions. The results of such a doctrinal development are likely to be very different from the rules that would be constructed in response only to a concern for protecting justifiable reliance on unilateral promises.

III. CONCLUSION: IMPLICATIONS OF THE CONTRAST

The two models of contract contrasted in this Article embody drastically conflicting visions of the relation between contract and community. The promise model envisions a marked antithesis between contractual relationships and community relationships. Contractual rights and obligations are rigidly defined by the promises of the contracting parties, and trust between the parties is confined

116. *See supra* notes 66-77 and accompanying text.

117. C. FRIED, *supra* note 1, at 37 (emphasis in original).

to the explicit terms of the promises. Structuring contract as unilateral obligations running in opposite directions focuses attention on the zero-sum aspects of exchange and thereby accentuates and justifies divisiveness and selfishness.[118] Emphasizing discreteness and abstraction further constricts personal involvement in the contractual relationship.[119] Community relationships would be disrupted by such a regime of rigidity, selfishness, and limited personal involvement, for "[c]ommunal life needs to maintain the lines of right and duty fluid in attention to an untrammeled trust. It must subordinate the jealous defense of individualistic prerogative to the promotion of shared purpose and the reinforcement of mutual involvement."[120] Accordingly, the promise model would maintain a sharp division between contractual relationships and community relationships.

In contrast, the exchange-relationship model seeks to heighten the interaction between contract and community. It underscores the flexibility and porosity of contractual obligation, stresses mutual dependence and cooperation between contracting parties, protects and encourages intense personal involvement. This model rejects the simplistic dichotomy of community as altruistic sentimentalism and contract as self-interested materialism.[121] It recognizes that contractual solidarity derives largely from the permeation of societal norms[122] and that social solidarity is enhanced by furthering contractual relationships.[123] It envisions contract and community intermingling, with each realm richer for its association with the other.

---

118. *See* I. MACNEIL, *supra* note 20, at 17-18.

119. Discrete transactions . . . are nonprimary relations. They involve only a small part of the personality, are very limited in scope, are nonunique in personal terms, and hence can be transferred readily. . . . The satisfactions derived are limited to the narrow economic exchange being accomplished. Buying gasoline for cash at a busy self-service station in a strange town is a fairly good modern example.

Id. at 13.

120. Unger, *supra* note 103, at 624.

121. *See* Olsen, *The Family and the Market: A Study of Ideology and Legal Reform*, 96 HARV. L. REV. 1497 (1983); Unger, *supra* note 103, at 641-42.

122. *See* I. MACNEIL, *supra* note 20, at 14, 58.

123. *See id.* at 93-94, 120 n.4.

Forthcoming in *The Handbook of Law & Economics*
(final draft: June 5, 2006)

Chapter on the Law & Economics of Contracts

Benjamin E. Hermalin, Avery W. Katz, and Richard Craswell

Contents

1 Introduction **1**
1.1 The economic motive for contracts 2
1.2 Law & economics issues in contracting 7
1.3 What this chapter is not 11
1.4 Organization of the chapter 12

2 Freedom of Contract **13**
2.1 Freedom of contract defined 13
2.2 The economic case for freedom of contract 16
2.3 The economic case against freedom of contract 24
2.4 Other arguments for regulating private contracts 40
2.5 Legal doctrines regulating freedom of contract 42

3 Formation of Contracts **52**
3.1 Pre-contractual behavior 52
3.2 Avoiding miscommunication 56
3.3 Legal doctrines addressing contract formation 57

4 Interpretation of Contracts: Contractual Incompleteness **63**
4.1 Modeling incomplete contracts 64
4.2 The sources of contractual incompleteness 70
4.3 Consequences of contractual incompleteness 76
4.4 Legal doctrines addressing contractual incompleteness 82
4.5 Overall assessment of the law of contract interpretation 94

5 Enforcement of Contracts **94**
5.1 General issues in enforcement 95
5.2 Monetary damages for breach of contract 97
5.3 Complications in determining monetary damages 110
5.4 Private enforcement of contracts 116
5.5 Other law bearing on contract enforcement 123

6 Conclusions **123**

1 Introduction

The essence of a free-market economy is the ability of private parties to enter into voluntary agreements that govern the economic exchange between them. Consequently, the law that governs such agreements is critical to the functioning of such economies. While the law of property determines the configuration of entitlements that form the basis of production and exchange, and the law of torts protects those entitlements from involuntary encroachment and expropriation, it is contract law that sets the rules for exchanging individual claims to entitlements and, thus, determines the extent to which society is able to enjoy the gains from trade. Accordingly, economists interested in the welfare properties of specific institutions in particular, or the micro-foundations of exchange generally, have good reason to take account of the law of contracts.

This chapter, accordingly, surveys the main issues arising in the economic analysis of contract law. We discuss both the main features of contract law as they relate to the problem of economic exchange, and how relevant legal rules and institutions can be analyzed from an economic perspective. In this introductory section, we set out the basic scope, methodology, and organization of the discussion to follow. Subsection 1.1 discusses why formal and informal contracts exist, and what economic functions they serve. Subsection 1.2 distinguishes between positive and normative issues in the economic analysis of contract law, and discusses some methodological problems associated with applying standard economic analysis to legal institutions and when engaging with legal scholarship. Subsection 1.3 identifies limits on the chapter's scope and provides bibliographic recommendations for material we don't cover; and subsection 1.4 sets out the organization of the remainder of the chapter.

A caveat is in order at the outset: although it is conventional to present contract law as a discrete field, one should understand that, to a significant extent, the operation of the rules and institutions discussed below will depend on other aspects of the law, including the fields of tort, bankruptcy, procedure, and evidence. Lawyers have a cliché that describes this interdependence; they say that "the law is a seamless web." It is useful to keep in mind that many issues that economists would regard as contractual, including some important limits on contractual freedom, are governed not by contract but by tort law. Additionally, the rules relating to certain categories of exchange, such as consumer, employment, insurance, and information-licensing contracts, have developed specialized content to the point that they are often treated as distinct legal fields. Finally, the practical ability of contracting parties to assert their legal entitlements depends importantly on the procedural rules that govern courts and other enforcement institutions. Many of the specific features of contract law that we discuss below cannot be understood except as a response to the costs and other limitations of such institutions.

		Seller: Buy insurance	Seller: Don't buy
Buyer	Buy insurance	$-p_b$, $-p_s$	$-p_b$, 0
	Don't buy	0, $-p_s$	$-\ell_b$, $-\ell_s$

Figure 1: Coordination of insurance payments

1.1 The economic motive for contracts

In a neoclassical exchange economy of the sort analyzed by Walras (1874) or Arrow-Debreu (Arrow and Debreu, 1954; Debreu, 1959), there is little need for contracts or contract law, since buyers and sellers can exploit all gains from trade through spot transactions. Indeed, in spot markets, such as public bazaars, the parties manage reasonably well without formal contracting. Contracting becomes worthwhile when there is a temporal element to their exchange or one party, at least, is unsure as to what her counterparty will do. For example, when the item to be exchanged needs to produced or the service being rendered takes time. Absent a contract, the parties could be reluctant to trust each other to complete the agreed upon exchange at the called-upon time, and thus valuable exchange is forgone. Conversely, contracts can be worthwhile even in non-exchange settings, as when advance commitment enhances the value of a gift by enabling reliance by the beneficiary (R. Posner, 1977, and Shavell, 1991) or when a supplier's commitment to remain in a market notwithstanding short-run losses deters competitive entry by rivals (*e.g.,* Rasmusen et al., 1991) or encourages entry by producers of complementary goods. The central question then becomes why commitment is valuable, to which there are several answers.

1.1.1 Coordination

The most straightforward reason to use contracts is to coordinate independent actions in situations of multiple equilibria. As an illustration, consider the game depicted in Figure 1, in which a buyer, b, and a seller, s, are independently deciding whether to purchase insurance against the loss of a good in transit. The parties' payoffs net of this decision are normalized to zero. Denote the insurance premium when purchased by party i by p_i and the *expected* loss suffered by party i when no insurance is purchased by ℓ_i. Assume that $\ell_b + \ell_s > p_b > p_s$; that is, going uninsured is more expensive than buying insurance, but it is cheaper for the seller to buy the insurance than for the buyer to do so. (This is perhaps because the seller, who packages the goods for shipment, is better able to control moral hazard and thus can obtain a better rate.)

If the parties make their choices independently, there are three Nash (1951) equilibria to this game: one in which the seller buys insurance, one in which the buyer buys insurance, and a mixed-strategy equilibrium in which both buy insurance with positive probability. Of these, the first equilibrium is the efficient one (and it is also more efficient than any disequilibrium outcome). A contract to play this efficient equilibrium could serve to ensure this outcome is achieved.

Formal contracts are of course not the only way for parties to coordinate among multiple equilibria. The efficiency of the seller-buys equilibrium could make it a focal point for the parties (Schelling, 1960). The literature on "cheap talk" (*e.g.,* Farrell, 1987a, 1993) suggests that such coordination can, in principle, also be achieved by having the parties announce their intentions in advance.

In actual institutional settings, however, contracts offer more stability than focal points or mere announcements. In particular, they provide permanent authoritative records that can be used by parties who suffer from imperfect recall or by those who need to delegate performance to their agents or successors. Note that when contracts are used for pure coordination purposes, they are self-enforcing in the sense that it is in each party's private interest *ex post* to comply with the chosen equilibrium. Hence, those designing such contracts can devote most of their attention to problems of formation and interpretation, and relatively little attention to problems of enforcement. This coordination function of contracts has been rather less discussed in the law and economics literature than the incentive mechanism functions discussed below, but it may be by far the most important purpose that contracts serve in practice. As Myerson (2004) suggests, coordination games could be the best models through which to understand legal institutions generally.

1.1.2 Implementing exchange over time

A second reason for using contracts is to implement exchanges that depend on future events. For instance, consider an insurance contract that covers a loss that occurs in state 1, but not in state 0. Under this contract, the insured pays a premium to the insurer, in exchange for a casualty payment received in state 1, but not in state 0.

In the standard model of risk allocation, goods are state-contingent commodities; for example, an apple in state 0 is considered a different good than an apple in state 1. Treating goods as state-contingent commodities has the advantage of allowing direct application of standard analyses of exchange, but has the disadvantage of abstracting from real institutional issues. In particular, exchange that is mutually beneficial *ex ante* may not be *mutually* beneficial *ex post*. In the insurance example, the insured will not wish to pay the premium *ex post* if state 0 is realized, and the insurer will not wish to make the casualty payment *ex post* if state 1 is realized. Such exchanges cannot, therefore, be implemented in spot markets and require some form of advance commitment.

In the typical insurance context, the transaction is motivated by the insured's risk aversion. But the need to contract across different states is more general than that and is not reliant on risk aversion. For instance, consider purely speculative exchange between risk-neutral parties, in which trade is motivated by differences of opinion regarding the probability of future events. If one trader thinks the price of orange juice will rise next year and another thinks it will fall,

they can make themselves better off *ex ante* by entering into a forward exchange in which the second promises to deliver to the first. As with the insurance contract, this exchange requires a commitment mechanism since *ex post* one of the parties is sure to regret the deal.

An analogous problem arises with the rental of capital assets or the extension of credit. Even though the owner of an asset may not be its highest-value, she may be unwilling to yield possession to a higher-value user for fear that she will be unable to get it back at the end of the rental period. The law of property provides a partial solution to this problem by entitling the owner to reclaim her asset, but evidentiary difficulties make this alternative an imperfect one (as illustrated by the old maxim, "possession is nine tenths of the law"). The party in possession could claim, for instance, that the transaction was a gift or a sale, or that the agreed lease period had not yet expired. In such settings, a contract that specifies the parties' relative rights and duties makes the borrower's promise to return the asset more credible, facilitating exchange.

More generally, some form of commitment is necessary in any exchange in which performance is sequential, because the party who performs first is effectively extending credit to the party who performs second. It may be possible to structure the exchange so that each stage of a party's performance is timed to coincide with the performance of her counterparty (*e.g.,* an installment sale of goods in which each shipment is delivered C.O.D.), but in many instances such timing may be infeasible or costly. For instance, consider a grocery that requires regular delivery of a perishable commodity such as milk. The costs of making and receiving payment (keeping cash on hand, updating accounts, preventing embezzlement, etc.) generate substantial scale economies if disbursements for multiple shipments are combined into a single monthly payment.

Contracts can also be useful in situations of hidden information. In Akerlof's (1970) lemons model, for example, adverse selection can prevent efficient exchange when the quality of the good to be traded is known to the seller but not to the buyer, even if the buyer values the good more. This problem can be overcome if the seller of a high-quality good can signal its quality by taking an action that is cheaper for her to take than it would be for the seller of a low-quality good. A common action in this regard is to offer a warranty against the good's proving to be substandard (Grossman, 1981). Conversely, the buyer could screen for quality by offering a premium to any seller who agrees to provide a warranty. The signal (screen) works only if the seller is bound to honor the warranty, because a low-quality seller can offer (agree to) a worthless warranty just as cheaply as a high-quality one can. Some form of commitment is, thus, needed to implement the exchange.

1.1.3 Implementing production over time

Finally, contracts are valuable in promoting production in advance of exchange. Advance production typically increases the surplus available from exchange, but requires sinking resources in ways that may be unrecoverable if the contemplated exchange is not completed. For example, a clothing manufacturer can increase

the price it receives for its products by producing them to meet the needs of its buyers, either very specifically (*e.g.,* custom-tailored suits) or only moderately so (*e.g.,* cutting them so they will be in style for a limited time only). Once the materials used to make the clothing are combined in a particular way, however, they can no longer be easily reconfigured to produce other items. In such settings, producers will be reluctant to sink such expenditures up front unless they can be assured that they will recover their costs *ex post* (see Williamson, 1975, for a seminal analysis of holdup problems).

As Katz (1996a) discusses, suppliers typically cannot capture all the surplus their upfront (pre-trade) investments generate; some of this surplus will go to the buyer. Absent binding purchase commitments prior to investment, suppliers' incentives to invest will be suboptimal, possibly to the point that no investment and, so, no trade occur. Binding contracts can restore proper incentives. Conversely, buyers can also increase their surplus from exchange by making up-front investments, whether out-of-pocket (*e.g.,* buying complementary inputs) or implicit (*e.g.,* ceasing to maintain alternate sources of supply). Because such investments are often relationship specific, however, buyers will not make them unless they can be assured that the exchange price will stay sufficiently low. In some cases, the parties may be able to provide such assurance by manipulating property rights (*e.g.,* Grossman and Hart, 1986; Hart, 1989) or industrial structure (*e.g.,* Shepard, 1987) *ex ante*, but when they cannot do so cheaply, contracts could be a cost-effective alternative.

1.1.4 Limitations of contracts as commitment devices

While contracts are often useful for achieving commitment, they can be imperfect devices for doing for some of the following reasons.

Specification costs. Because it is costly to foresee or to write down all the potential contingencies that might be relevant to the performance of the parties' contractual obligations, actual contracts are often left incomplete. Incompleteness has at least two meanings: first, the contract could simply fail to provide for certain contingencies, in which case a tribunal called upon to enforce the contract, or the parties themselves, would have to decide after the fact what to do if such contingencies arise. Second, the contract could cover all relevant contingencies, but not in as fine-tuned a manner as would be ideal insofar as the contract does not distinguish finely enough, in terms of consequent obligations, among the possible contingencies. In either event, the contract will, with positive probability, fail to assure commitment or commit the parties to a course of action that is suboptimal *ex post*.

Enforcement costs. It is never costless to hold a party to his commitment if he is inclined to try to escape it. If the contract is being enforced through the courts, for instance, lawyers must be hired and evidence assembled, and performance or damages are likely to be awarded only after some delay. Such costs make enforcement incredible when the damages from breach are relatively

small and parties can exploit this lack of credibility by holding the level of breach below the threshold necessary to provoke suit (Menell, 1983; Priest, 1978).

Unobservable and unverifiable actions. Even if legal commitment has been established and the means for its enforcement are available, the beneficiary of a contractual promise may be unable to determine whether the promise has been kept or broken. For instance, the typical purchaser of a complex consumer product is not in a position to tell whether the product has been manufactured according to warranted specifications; at most she can observe whether the product works as she expected. Even if a promisee can determine that there has been a breach, she may nevertheless be unable to demonstrate that fact to a third-party enforcer at reasonable cost. For instance, a supplier might deliver substitute goods that appear reasonably equivalent to a lay person or to a generalist court, but which the parties themselves know to be substandard. In such situations, the promisee's inability to prove that the promise has been breached renders it ineffective as a method for assuring commitment. On the other hand, as we discuss in §4.3.1, the parties can sometimes contract around the court's lack of expertise (see, *e.g.,* Hermalin and Katz, 1991; Maskin and Tirole, 1999).

Dynamic inconsistency. In cases where the purpose of contractual commitment is to promote specific investment, the parties' incentives to stick with their deal may change after the investment has been completed (*e.g.,* Laffont and Tirole, 1988; Aghion et al., 1994). In particular, the parties may all wish to modify or renegotiate their bargain. But if the parties anticipate that renegotiation will take place, it could prove impossible to induce them to undertake efficient investments *ex ante*.

The need for pre-contractual commitment. Some commitments, in order to serve their purpose, must be undertaken before the parties are in a position to engage in voluntary contracting. For example, parties may spend resources on finding contractual partners or on determining whether exchange is worthwhile. Even once an available partner and potential transaction are identified, it typically takes time and expense to negotiate terms; and commitments are often less valuable if they are delayed until bargaining is completed. (For an extreme example, consider the case of an emergency paramedic who must decide whether and how to treat an unconscious accident victim who is not carrying an insurance card.) In ongoing or repeated relationships it is possible for the parties to agree to accept liability in advance of a final bargain, but in one-shot or new relationships it is not.

The law of contracts has recognized most of these problems and has devised a variety of doctrinal arrangements to deal with them; and the succeeding sections of this chapter will discuss such arrangements in more detail. The reader should appreciate at the outset, however, that because these legal arrangements are themselves imperfect, parties will often want to use legal contracts in combination with other legal and non-legal commitment devices, such as deposits, third-party guaranties, reputational bonds, repeated dealing, mutual threats, hostage exchange, investing in altruistic preferences, and the like. The success of formal contract law, accordingly, depends importantly on how well it functions in combination with these substitute and complementary devices, and not just on how well it works in isolation.

1.2 Law & economics issues in contracting

1.2.1 Normative issues

Much normative discussion relating to contract law revolves around the issue of freedom of contract—to what extent will unregulated private contracting lead to desirable social consequences? We discuss this issue, and its relationship to standard issues in welfare economics, in §2 below. At the outset, however, it is worth observing that the dominant normative consideration here, even more so than in other fields of law and economics, is transactional efficiency. In part, this dominance follows the implicit assumption, shared by most commentators, that externalities and analogous market failures are a less significant phenomenon in this field of law than they are in, say, tort law. But the focus on efficiency also stems from the general recognition that much of contract law, putting aside specialized categories such as consumer and employment contracts, is designed for the purpose of facilitating exchange between business firms and analogous commercial entities. Such entities are motivated primarily by economic gain as opposed to nonpecuniary considerations, enter into legal obligations deliberately and at arms length, and are rational in the standard economic sense. It is thus easier to justify applying the efficiency norm to such voluntary arrangements than to the typical tort case involving persons drawn together involuntarily and outside of market institutions.

A more complete account of social welfare, however, would consider competing normative values such as fairness, equity, etc. to the extent they affect social well-being. While there has been relatively little economic analysis of contract law in this regard, we will discuss these values insofar as they are relevant to specific analytical and doctrinal topics.

1.2.2 Positive issues

While most work on the economics of contract law has sought, at least in part, normative conclusions, there is a segment of the literature devoted to predicting and explaining how different contractual rules affect private transactions, and why contracting parties might choose one contractual device rather than others. For example, a variety of authors (*e.g.,* Joskow, 1987; Crocker and Masten, 1988; Pirrong, 1993) have investigated the connection between the use and duration of contractual agreements and the extent of relationship-specific investments. Other authors (*e.g.,* Klein, 1980; Goldberg and Erickson, 1987; Hadfield, 1990; Gergen, 1992) have sought to explain the common use of indefinite or open terms in otherwise clearly negotiated agreements; and still others

(*e.g.,* Weinstein, 1998; Goldberg, 1998, 2000) have sought to explain particular risk-sharing or option terms. In the field of commercial contracts, there is a vigorous literature discussing the determinants of secured lending (see, *e.g.,* Scott, 1986; Schwartz, 1989; Triantis, 1992; Mann, 1997a,b). The antitrust literature has considered whether certain contractual practices are more likely to have efficiency or anticompetitive motivations (*e.g.,* Cheung, 1969; Kenney and Klein, 1983, 2000; Crocker and Masten, 1988; Klein and Murphy, 1988; Masten and Snyder, 1993). While a full survey of this literature is beyond the scope of this chapter, the reader should be aware that many of the issues discussed in the later sections have been discussed empirically. At the same time, however, it also true that the empirical study of contract is relatively less developed than the theory, leaving much room for future researchers.

1.2.3 Economic versus non-economic theories of contract law

In recent years, the majority of contracts scholarship in the legal academy has reflected a methodology based on economic analysis, and most legal scholars in the field have become conversant with economic concepts such as efficiency, moral hazard, adverse selection, and the like. (Conversely, over the same period, economic theorists have increasingly come to appreciate the importance of legal concepts such as principal and agent.) Nonetheless, a considerable amount of discussion in legal circles continues to reflect alternative conceptual frameworks; and economists engaged in interdisciplinary work should be aware of these competing frameworks and their underlying assumptions. Three major competing perspectives are worth brief discussion here; we denote these as the corrective justice, liberal autonomy, and social constructivist perspectives.

Corrective justice. Corrective justice, the most longstanding of these perspectives, has intellectual roots that trace back to classical writers such as Aristotle. This perspective holds that judicial institutions are only justified in acting to redress unjust or wrongful situations. Examples of such redress in the contractual setting would include restitution of unjustly received benefits or compensation for wasted expenditures incurred in reliance on a broken promise.

The corrective justice approach seeks the restoration of some past or proper state of affairs, and thus can stand at odds with the economic approach to law, which tends to regard past gains and losses as sunk and to emphasize incentives for future behavior (*e.g.,* Easterbrook, 1984). It is less clear, however, that the corrective justice approach has any implications at all for *ex ante* analysis of legal problems; and many legal writers in this tradition (*e.g.,* Dworkin, 1980) have distinguished between the use of economics in judicial settings, which they regard as requiring decisions according to principle, and its use in legislation and contractual planning, which may legitimately be designed to promote goals of social policy or private advantage.

Liberal autonomy. The liberal or autonomy-based perspective (see, *e.g.,* Barnett, 1986) emphasizes the individual as opposed to the collective interest.

From this point of view, individual rights should take priority over more general concerns of society. Such a perspective is plainly more consonant with the economic approach than is the corrective justice perspective. Standard economic measures of welfare are based on aggregates of individual utility, and under conventional assumptions of welfare economics, liberal freedoms tend to lead to desirable economic outcomes (see §2.2 below). There will be tension between the liberal and economic approaches, however, whenever market failures or transaction costs prevent the completion of efficient exchanges, as Sen (1970) has shown through social choice theory. In a contractual setting, for instance, liberal theorists might argue that obligations to which the parties have not knowingly consented may not be imposed on them even when those obligations would be both efficient and distributionally equitable. Conversely, libertarians might argue that the law should protect even anticompetitive agreements, such as price-fixing, on the basis that the conduct is voluntary and that consumers have no inherent right to trade with producers on any particular terms.

Social constructivism. The social constructivist perspective takes the position that the normative goals of society—and in the view of many writers, its descriptive categories as well—are determined by collective and ongoing deliberation among its citizens, and accordingly, that the main goal of legal institutions should be to provide adequate opportunity for such decision making. So defined, this perspective can be seen to encompass a variety of more specific normative positions, including those of writers who emphasize civic virtue (Kronman, 1987), democratic self-government (Pildes and Anderson, 1990), an egalitarian distribution of political and economic power (Kennedy, 1982), or the primacy of particular substantive values, which economists would call merit goods (Radin, 1987). On such views, the constituted citizenry may well decide to pursue economic goals such as efficiency, but may, with equal legitimacy, decide to pursue other procedural or substantive goals. For example, if the citizenry decided the well-being of local manufacturers were sufficiently important to outweigh productive efficiency or economic liberty, this decision would legitimize restricting interregional trade. It follows that there is no particular reason why individualist institutions, like contract or the market, are more legitimate venues for such decision making than collective or political ones.

Relationship between economic and non-economic theories. These three rival perspectives can and often do combine to provide overlapping arguments against the use of economic analysis in contract law. For example, the most prominent alternative theory of contract law to be put forward in recent years, the so-called "will theory" (*e.g.,* Fried, 1981), holds that promises ought to be kept for their own sake, in part because promise-breaking is a deontological wrong that needs to be rectified as a matter of corrective justice, and in part because enforcing promises is necessary to respect the autonomy of the promisee (and arguably the promisor as well). In response to such arguments, more economically oriented writers (*e.g.,* Craswell, 1989a; Shavell, 1991) have

replied that most economic analysis of contract law is aimed at filling the gaps in incomplete agreements and setting default rules that operate when the parties have expressed no preference regarding a particular issue. If we accept that most business contracts have economic purposes, economic analysis will help those interpreting contracts to better implement the parties' will.

The will theorists might offer, as a rejoinder, the observation that there is a difference between what most contracting parties subjectively understand when they make and receive promises, and what would be most efficient for them to do. For instance, most people, even those with substantial business experience, intuitively understand promises to bear inherent moral force and believe that the mere fact that it turns out to be suboptimal to carry them out should not count as an excuse for non-performance. If this claim is true—and it should be plain that it is an empirical claim—then will theorists (*e.g.,* Charny, 1991) would argue that it violates the parties' autonomy and dignity for the state to enforce the efficient bargain, which they perhaps should have made, rather than the actual arrangements that they did make.

An obvious response for economists to such non-economic considerations is to incorporate them into a larger welfare analysis in which the relevant non-economic values are interpreted as arguments to be traded off against one another in a Bergson-type social welfare function (SWF). Shavell and Kaplow (2002) have presented such an analysis at length, but have achieved mixed success in persuading non-economically-oriented legal scholars of its merits. Observe, in this regard, that the three main categories of non-economic theories vary in their compatibility with the SWF approach. Pure corrective-justice theorists would reject the SWF on the grounds that legally appropriate actions can only be justified with reference to right and wrong, and never with reference to consequences. Liberal autonomy theorists tend to be somewhat more accepting of the SWF approach, but they would reject the idea that rights could be traded off against each other or against economic values (formally, this is equivalent to adopting a lexicographical preference ordering for the SWF). Social constructivists would view the whole SWF approach as logically circular, since they view the social values that would underlie any SWF as endogenously determined by political and cultural processes of which economic policy discussions are an integral part. In their view, starting with a SWF and then attempting to maximize it is putting the cart before the horse.

Most importantly, economists who study legal institutions should recognize that lawyers and legal commentators often employ different methodologies than they do. In particular, legal scholars do not typically draw the same sharp distinction between positive and normative analysis as economists do. On the contrary, many seemingly descriptive statements made by lawyers or appearing in legal texts are understood in context to carry significant normative overtones, and *vice versa*. Non-lawyers who are insufficiently appreciative of the mixed nature of such discourse may miss an important part of what is being said, and may be led to model the phenomena under discussion in an incorrect or misleading way (Katz, 1996a). Similarly, lawyers' customary method of reasoning inductively from individual cases, rather than deductively from general principles,

makes many lawyers reluctant to accept some of the standard methodological practices of economists, including formal modeling and the use of statistical aggregation. When presenting positive arguments or analyses to legal audiences, accordingly, it is generally necessary to lay out one's methodological assumptions explicitly and at the outset. Otherwise, one risks being misunderstood by those lawyers who are accustomed to assuming a normative subtext or an individualist perspective whenever legal topics are being discussed.

1.3 What this chapter is not

1.3.1 A guide to contract theory

Not surprisingly, there is a strong link between the law and economics of contracts and the economic sub-discipline of contract theory.[1] Contract theory provides a framework to analyze the scope and limits of what contracts can accomplish, at least at a theoretical level. It is, however, beyond the scope of this chapter to offer a comprehensive guide to contract theory. To be sure, some contract theory will be discussed as appropriate within the context of the issues discussed below; but no one should mistake such a scatter-shot approach for an attempt at a systematic treatment of the subject.

For readers interested in contract theory, there are a number of excellent introductions. Most new graduate microeconomics texts devote at least a chapter to the subject (see, *e.g.,* Kreps, 1990; Mas-Colell et al., 1995; Varian, 1992). Good book-length introductions are Laffont and Martimont (2002) and Bolton and Dewatripont (2005). For those looking for cheap—in fact free—introductions, one of us (Hermalin) hosts two introductory manuscripts (Caillaud and Hermalin, 2000a,b) on his web page.

1.3.2 A guide to the law of contracts

While this chapter discusses most of the important economic issues relating to the law of contracts, it does not attempt to present a survey of major legal doctrines in the field. It is worth noting, as a general observation, that many legal doctrines may seem indistinguishable from an economic viewpoint. Nevertheless, they vary in their specific content and their differences are of practical importance to lawyers and clients. For example, the doctrines of mistake, impossibility, and frustration of purpose all excuse contractual liability in extreme or unexpected circumstances and thus allocate risk to the recipient of a contractual promise, but their legal application varies.

Economists interacting with lawyers or pursuing research in contract law, accordingly, should be aware of distinctions such as these and of the associated views of major legal authorities. The most useful general treatise on US contract law is probably Farnsworth (2004), available in both one- and three-volume editions; a shorter introduction to the subject can be found in Chirelstein (2006).

[1] Whether this link is a valuable one is, however, a matter of debate. See, for instance, the *Yale Law Journal* debate among E. Posner (2003a), Ayres (2003), and Craswell (2003).

For statutory material applicable to US contracts for the sale of goods, the best source is White and Summers (2000).

Less attention has been paid by economists to non-US legal systems, making them a potentially fruitful source for future research; here leading English-language treatises include Honnold (1999) and Schlechtreim (1998) on international sales contracts, Bonell (1997) on other international commercial contracts, Atiyah (2003) and Treitel, ed. (2003) on English contract law, and Lando and Beale, eds. (2002) on the contract law of the civil law systems of continental Europe. In this chapter, however, most discussions of legal doctrine will be restricted to US law.

1.4 Organization of the chapter

The remainder of this chapter is divided into four sections corresponding to the major conceptual divisions of contract doctrine. Section 2, entitled "Freedom of Contract," discusses the scope of private parties' power to create binding contractual obligations. It analyzes the major doctrines that govern which bargains will be recognized and enforced by state legal institutions, and which parties are empowered to create enforceable bargains. It also considers how doctrinal limits on freedom of contract correspond to the economic criteria for determining whether decentralized trade will lead to optimal welfare outcomes.

Section 3, entitled "Formation of Contracts," discusses the extensive body of legal doctrine that governs the procedural mechanics of exchange, as well as the rules that govern the parties' obligations before they enter into exchange. These formal rules, by attaching consequences to the various acts and omissions that bargainers can choose from in searching for and negotiating with potential contractual partners, affect parties' incentives to make and to respond to offers, to delay, to bluff, and to communicate with one another in the first place.

Section 4, entitled "Interpretation of Contracts: Contractual Incompleteness," discusses the problems that arise when it is unclear whether those parties who are empowered to create binding contracts have actually done so, and if they have, what specific obligations they have created. Recent work in microeconomic theory has also been concerned with this problem, especially as it relates to the ability of third-party enforcers to verify the parties' bargain. From a legal perspective, the problem is governed by the various doctrines dealing with contract interpretation, and this section shows how the legal rules in this area affect and respond to the economic problem of incomplete contracts.

Section 5, entitled "Enforcement of Contracts," discusses how the foregoing rules and institutions are translated into effective costs and benefits that can motivate parties to comply with their obligations and to insure against others' lack of compliance.

Finally, the concluding section offers some overall perspectives on the entire discussion, relates its main points to analogous or complementary doctrines in related fields of law, and offers some speculations regarding the path of future legal and economic developments in the area.

2 Freedom of Contract

The threshold issue in any discussion of contract law is freedom of contract—the extent to which the law sanctions the use of contracts as a commitment device. No legal system enforces all voluntary private agreements, but in the US and other industrial democracies, most contracts that support legitimate economic exchange are at least presumptively enforceable. Still, the limits of freedom of contract vary among Western countries and are an important element of regulatory policy. This section, accordingly, analyzes and evaluates those limits in economic terms. Subsection 2.1 defines the scope of the issues included under the framework of contractual freedom; subsection 2.2 reviews the presumptive economic case in favor of freedom of contract; subsections 2.3 and 2.4 discusses the main arguments, economic and otherwise, that are typically used to justify limits on private contracting; and subsection 2.5 outlines the major doctrinal limitations on freedom of contract that are in force in the US and in related systems, and relates those doctrines to the economic arguments set out in the prior subsections.

2.1 Freedom of contract defined

2.1.1 State regulation versus state enforcement

The concept of contractual freedom encompasses a number of distinct considerations. One important distinction is between negative and affirmative government sanction: are the parties permitted to enter into a given contract versus will the law enforce it? These two questions are not equivalent: there are many agreements that cannot be enforced in the courts but that can still be useful as commitment devices if the parties can manage to implement them privately. For instance, prior to the passage of the Sherman Act, contracts in restraint of trade were unenforceable under US common law; this reduced their incidence, but not to the point of elimination. Under modern antitrust statutes, in contrast, government disapproval of anticompetitive conduct goes beyond non-enforcement to include active interference through civil liability and, in some cases, criminal prosecution.

In the remainder of this section, accordingly, we focus on those limits on contractual power that are motivated by regulatory concerns that the agreement itself is socially undesirable for reasons of inefficiency, inequity, and other substantive objections.

2.1.2 Positive versus negative contractual freedom

The freedom to enter into contractual liability would be rather less meaningful were it not accompanied by the complementary freedom to avoid liability for contracts into which one does not wish to enter. In general, this negative freedom applies to most types of contractual obligations, but not all. In traditional common law, for instance, some businesses (*e.g.,* mills, ferryboats, railroads, and the like) are designated as common carriers and are obliged to enter into

exchange on standard terms with anyone who wishes. Modern statutes have expanded such duties in a variety of ways: for instance, the essential-facilities doctrine in antitrust law requires vertically integrated firms to make certain stages of production available on a contractual basis to their non-integrated competitors; and anti-discrimination laws require businesses to deal with customers and suppliers on an equal basis without regard to race, religion, or the like.

Additionally, some rules of tort and property law have the effect of requiring rightholders, under some circumstances, to transfer their entitlements to persons in need or to the general public. For instance, the doctrine of eminent domain requires landowners to convey their property to the state when required for an appropriately defined public use. (As in the US, the government could be constitutionally obliged to pay just compensation when it takes private property.) Similarly, the tort doctrine of necessity allows parties in dire need to make use of others' property when voluntary contracting is not feasible (as with an endangered hiker who breaks into an empty cabin to find food or shelter) so long as compensation is paid *ex post*.[2] Such doctrines are typically conceptualized by legal theorists as internal limits on the underlying entitlement at issue rather than as restrictions on contractual freedom; however, they usually amount in practice to restrictions on contractual freedom because of the infeasibility of contracting around them. One could theoretically imagine contracting in advance with the government not to exercise its right of eminent domain, but such contracts are rare and it is doubtful whether they are immune to abrogation by subsequent governments.

Finally, some doctrines of contract law impose promissory liability even when the promisor has not actually intended to enter into a contractual exchange. The doctrine of promissory estoppel, for instance, holds parties to promises on which a reasonable person would foreseeably rely, at least to the extent necessary to protect the promisee's reliance; and the doctrine of trade usage holds parties to contracts that experienced market participants would view as legally enforceable under similar circumstances, even if the parties had not themselves subjectively intended to be bound in the particular instance. For the most part, such doctrines provide rules of interpretation rather than of substantive contractual freedom, in that it is possible to avoid liability by being sufficiently explicit in one's communications. In some cases, however, the law does not allow one to disclaim liability for one's representations or promises; for instance, commercially sophisticated parties dealing with less sophisticated counterparts can find themselves bound to statements made to the less experienced party even if sophisticated persons would understand such statements do not entail legal obligation. Such restrictions on disclaimers are usually motivated by considerations of market failure, transaction cost, or distribution, as discussed below.

[2]Modern legal theorists, following Calabresi and Melamed (1972), use the concept of property and liability rules, further discussed in this *Handbook*, to identify situations in which such involuntary exchanges are authorized by law: entitlements subject to such imposition are said to be protected by a liability rule, while entitlements that are immune from such imposition are said to be protected by a property rule.

2.1.3 Mandatory versus default terms

Finally, as the example of promissory estoppel illustrates, it is useful to distinguish between mandatory contract terms, which the parties are not legally free to change, and default terms, which the parties are theoretically free to change but which govern the contract to the extent the parties are silent. An example of the former would be the constitutional prohibition on involuntary servitude, which, for instance, prevents people from pledging their future labor for long periods, even when they might find wish to do so, such as to provide collateral for a loan. An example of the latter would be the various warranties of quality that are implied under current US law in contracts for the sale of goods. Sellers are generally permitted to disclaim such warranties, subject to some limits imposed by consumer-protection and product-liability law, by following the requisite procedures, which usually require some specific notification to the buyer.

Because it is costly to write complete contracts, all systems of contract law must provide default terms to cover the issues over which the parties do not specifically bargain (see §4.4.1 below). The regulatory effects of default terms, however, are bounded by the costs of contracting around them. It could be a reasonable approximation to ignore such effects in many instances, especially in the commercial setting where parties are sophisticated and have access to legal advice. In some cases, however, re-contracting costs are substantial and the choice of default rule will have the effect of privileging one outcome over others. For example, in mass transactions in which parties communicate through standard forms, it is impractical to reconcile all the discrepancies between the various forms; and under modern doctrine, the legal default applies to all issues on which the forms do not agree. This often has the consequence of providing broad product warranties and leaving the seller open to liability for damages following breach, even if the seller attempted to disclaim such liability.

As with the decision to withhold enforcement, the provision of default rules may be motivated either by regulatory purposes or by the desire to conserve on transaction costs. In situations where there is some doubt about whether a specific contract term should be discouraged, for example, a default rule supplies some deterrent effect while still allowing parties whose gain from the term is sufficiently high to opt out of the default at a price. For example, Camerer et al. (2003) advocate using default rules as a relatively libertarian method of regulating against poor decisions caused by bounded rationality. Or, as Ayres and Gertner (1989, 1992) have suggested, and as we discuss further in §4.4.1 below, a default rule may be employed as a screen to induce parties to reveal private information that might be relevant to *ex post* interpretation or to the *ex ante* decision whether to enter into exchange.

2.2 The economic case for freedom of contract

2.2.1 Welfare economics

An economic case for or against freedom of contract is based on the consequent welfare implications. In this section, we briefly review what a welfare analysis of markets suggests about freedom of contract.

Economists typically use two welfare criteria. One, known as Pareto efficiency, evaluates a proposed allocation among a set of actors by asking whether there exists a second allocation that (i) none of the actors prefer less than the proposed allocation and (ii) at least one of the actors actually prefers to the proposed allocation. If such a second allocation exists, the proposed allocation is deemed inefficient (alternatively, Pareto inferior or Pareto dominated). The second allocation in this case is deemed Pareto superior. If no such second allocation exists, the proposed allocation is deemed efficient.

While the Pareto criterion is useful for ruling *out* undesirable allocations, it doesn't always serve as a useful guide for selecting a desirable allocation. For instance, it offers no guidance as to who should receive an indivisible object; any allocation other than throwing it away is Pareto efficient because a switch to another allocation would not be favored by the party losing the object.

An alternative welfare measure is to consider a function that aggregates, in some way, the preferences of the actors in question. A full discussion of welfare functions is beyond the scope of this chapter.[3] We will limit attention to the utilitarian welfare function, $W = \sum_{i\in\mathcal{I}} u_i$, where $\mathcal{I}$ is an index set over the actors and u_i is the utility the ith actor enjoys from the proposed allocation (if there is a stochastic aspect to the allocation, u_i should be understood to be i's *expected* utility). Any allocation that maximizes social welfare, W, must be Pareto efficient;[4] but Pareto efficiency does not necessarily imply that social welfare is maximized. For instance, as noted, any allocation of an indivisible object (other than throwing it out) is Pareto efficient, but only the allocation that awards it to the person who values it most is welfare maximizing.

A stronger connection between the two welfare criteria can be achieved if one accepts the existence of a transferable good (typically taken to be money). Now the "losers" from moving to a social welfare-maximizing allocation can receive payments from the "winners" as compensation. If preferences can be captured by a quasi-linear utility function (*i.e.,* of the form $u + y$, where y is money and u is utility from other goods), then an allocation is Pareto efficient

[3]The interested reader is directed to Chapter 22 of Mas-Colell et al. (1995) or Chapter 6 of Laffont (1988). Also see Arrow's (1963) classic book.

[4]Proof: Suppose not. Then, although the allocation $\mathbf{a}$ maximizes W, there is another allocation $\tilde{\mathbf{a}}$ that Pareto dominates $\mathbf{a}$. Let u_i and $\tilde{u}_i$ denote the utilities under the allocations $\mathbf{a}$ and $\tilde{\mathbf{a}}$ respectively. By the definition of Pareto dominance, $\tilde{u}_i \geq u_i$ for all i and there is at least one i such that $\tilde{u}_i > u_i$. But, then,

$$\sum_{i\in\mathcal{I}} \tilde{u}_i > \sum_{i\in\mathcal{I}} u_i \,,$$

which contradicts the assertion that $\mathbf{a}$ maximizes social welfare.

if and only if it maximizes welfare.[5] For this reason, economists are generally satisfied with social welfare (total surplus) as an appropriate welfare standard when transfers are feasible.[6] Note that this analysis relies on the marginal utility of the transferable good being constant across individuals, so that we are still seeking to maximize $\sum_{i\in\mathcal{I}} u_i$; transfers among the actors are irrelevant to maximizing welfare because the benefit one actor gets from receiving a dollar is completely offset by the cost another incurs transferring that dollar.

Competitive markets are typically seen as doing well with respect to the Pareto efficiency criterion. Under somewhat stringent conditions—in particular, (i) that complete markets for all commodities (including commodities such as clean air and water) exist, (ii) that no actor has market power (*i.e.,* acts as a price setter rather than a price taker), and (iii) symmetric information—a general equilibrium of the economy will be Pareto efficient (see, *e.g.,* §6.3 of Debreu, 1959). This result is known as the First Welfare Theorem.[7]

The First Welfare Theorem applies to the economy as a whole. If the entire economy is an Arrow-Debreu economy and in equilibrium, then any particular market within the economy must be "efficient," insofar as any change in it

[5]Proof: Footnote 4 *supra* established that welfare maximization implies Pareto optimality. Consider a Pareto optimal allocation of real goods and money. Let u_i and y_i denote the utility components under this Pareto optimal allocation. Suppose this allocation does not maximize welfare. Then, there exists another allocation of real goods, with utility components u_i^* such that

$$\sum_{i\in\mathcal{I}} u_i^* > \sum_{i\in\mathcal{I}} u_i \,.$$

Because the "u_i" allocation is Pareto efficient, it cannot be that $u_i^* \geq u_i$ for all i; that is, a change in allocation must create "losers," for whom $u_i < u_i^*$. Let $\mathcal{L}$ be the set of losers and L be the number of elements in $\mathcal{L}$. As just noted, $L > 0$. Similarly, there must be winners (*i.e.,* those for whom $u_i^* \geq u_i$). Let $\mathcal{W}$ be the set of winners. For losers, define $\tau_i = u_i - u_i^*$. For winners, define $t_i = u_i^* - u_i$. Because the "u_i^*" allocation is welfare maximizing,

$$\sum_{i\in\mathcal{W}} t_i - \sum_{i\in\mathcal{L}} \tau_i \equiv G > 0 \,.$$

Finally, consider the allocation of real goods that produces utilities u_i^*, allocates $y_i - t_i$ in money to each winner and $y_i + \tau_i + G/L$ to each loser (note the additional transfers sum to zero, hence are feasible). For a winner, $u_i^* + y_i - t_i = u_i + y_i$, so winners are indifferent. For a loser, $u_i^* + y_i + \tau_i + G/L = u_i + y_i + G/L$, so losers are strictly better off. But, then, we have an allocation (including money) that Pareto dominates our original allocation, contradicting its Pareto optimality. The result follows by contradiction.

[6]The Kaldor-Hicks criterion is even more flexible: An allocation, $\boldsymbol{\alpha}$, over real goods is a Kaldor-Hicks efficient allocation if there is no other allocation over real goods $\boldsymbol{\alpha}'$ and no profile of transfers, $\mathbf{t}$, such that the overall allocation $(\boldsymbol{\alpha}', \mathbf{t})$ Pareto dominates the overall allocation $(\boldsymbol{\alpha}, \mathbf{0})$, where $\mathbf{0}$ means no transfers are made. It is sufficient that $\mathbf{t}$ exist—whether or not these transfers are made—for $\boldsymbol{\alpha}'$ to dominate $\boldsymbol{\alpha}$ according to the Kaldor-Hicks criterion. If all individuals have utility $u_i + y_i$, where u_i is an individual's utility over real goods and y_i is his allocation of money (transfer), then an allocation of real goods is Kaldor-Hicks efficient if and only if it maximizes $\sum_{i\in\mathcal{I}} u_i$. See Chapter IV of Arrow (1963) for details.

[7]There is also a second Welfare Theorem that has to do with the ability of prices to serve as appropriate incentive devices in a competitive (Arrow-Debreu) economy; that is, any Pareto efficient allocation can be supported as a general equilibrium of the economy by selecting the appropriate prices (see §6.4 of Debreu for details).

changes at least part of the overall allocation and no other overall allocation is Pareto superior. If, however, one doubts the Arrow-Debreu model adequately models the *entire* economy, one can still ask, in certain circumstances, whether a particular market in it is achieving an efficient allocation.

A single competitive market (*i.e.,* a market in which neither buyers nor sellers exercise market power) will achieve an equilibrium at the price that equates demand and supply. Under normal assumptions, this equilibrium is unique. Assume the demand curve is a good approximation for the social marginal benefit curve (*i.e.,* there are essentially no positive externalities from the good in question and income effects are *de minimis*).[8] Assume, further, that the supply curve is a good approximation for the social marginal cost curve (*i.e.,* there are essentially no negative externalities from the good in question). Then, as is well known, welfare is the area beneath demand and above supply from 0 to the number of units traded. This area—that is, total welfare—is maximized by exchanging the quantity that corresponds to the intersection of supply and demand. But this quantity is precisely the quantity that will be exchanged in competitive equilibrium—the competitive equilibrium maximizes total welfare and, thus, achieves a Pareto-efficient allocation.

2.2.2 Theoretical justifications for freedom of contract

The welfare-theoretic arguments of the previous subsection rely on the assumption of competitive markets. Many writers, however, have had the intuition that freedom of contract is desirable much more generally.

In the law and economics literature, this intuition is most prominently associated with the work of Ronald Coase and his widely-cited "Coase Theorem" (Coase, 1960). Despite its formal sounding name, the Coase theorem is not a theorem in the traditional sense (nor did Coase suggest it was). Indeed, as Medema and Zerbe (2000) point out, there is not even an agreed upon statement of it. We offer the following version:

Theorem 1 (Coase theorem) *Consider a bilateral contracting situation in which (i) the parties are rational with respect to their individual self-interests; in which (ii) the parties can agree on any contract without incurring transaction costs; and in which (iii) the parties' utilities are additively separable over the allocation of real goods and monetary transfers (i.e., are of the form $u_i + y_i$, where u_i is party i's utility from the allocation of the real goods and y_i is his or her net transfer). Then the allocation of real goods after contracting will maximize total welfare regardless of the initial allocation of real goods.*

In this formulation, the Coase theorem is a true theorem with the following proof: Let $\boldsymbol{\alpha}$ be any real-good allocation that does *not* maximize welfare. We need to show that no such $\boldsymbol{\alpha}$ will be implemented via contracting. Let $\boldsymbol{\alpha}^*$ be a welfare-maximizing allocation. Clearly, if $u_i(\boldsymbol{\alpha}^*) > u_i(\boldsymbol{\alpha})$ for both parties

[8] For a detailed discussion of measuring consumer benefit and the consequence of income effects for such measurements see Chapter 10 of Varian (1992). Also see Willig (1976).

i, then $\boldsymbol{\alpha}$ will not be implemented—because both parties act rationally and they can costlessly contract, they won't settle for $\boldsymbol{\alpha}$. Consider the only other possibility: $u_i(\boldsymbol{\alpha}^*) > u_i(\boldsymbol{\alpha})$ for one party and $u_j(\boldsymbol{\alpha}^*) \leq u_j(\boldsymbol{\alpha})$ for the other. Let $\tau_M = u_i(\boldsymbol{\alpha}^*) - u_i(\boldsymbol{\alpha})$ and $\tau_m = u_j(\boldsymbol{\alpha}) - u_j(\boldsymbol{\alpha}^*)$. The optimality of $\boldsymbol{\alpha}^*$ entails $\tau_M > \tau_m$. Pick any transfer τ such that $\tau_m < \tau < \tau_M$. Then a contract that selected allocation $\boldsymbol{\alpha}^*$ and had i transfer an additional τ to j would be preferred by both parties to a contract that implemented allocation $\boldsymbol{\alpha}$. Costless contracting and the parties' rationality thus rule out $\boldsymbol{\alpha}$ being the contracted-for allocation.

Corollary 1 *Under the assumptions of the Coase Theorem, interference or restrictions on the contract the parties sign cannot increase total welfare; that is, under these assumptions, there should be freedom of contract if the only welfare issue is the total welfare of the parties to the contract.*

How strong a case the Coase Theorem makes for freedom of contract depends on the appeal of its assumptions. Consider, first, assumption (iii); that the parties have quasi-linear utility functions. That assumption is common to most welfare analyses of partial-equilibrium settings. Although the assumption can frequently be justified, it is not always.[9] Suppose, for instance, that one party's utility is $u + y$ if $y \geq 0$, but $-\infty$ if $y < 0$. An interpretation is that this party simply cannot survive if made to make transfers. Now the Coase result could fail to hold if this party is the one who must transfer to ensure a welfare-maximizing outcome. On the other hand, if this party is the recipient of transfers, then welfare-maximization would continue to hold.

As Medema and Zerbe (2000) note, the Coase Theorem has two conclusions. One is an efficiency conclusion—private contracting will lead to a welfare-maximizing solution. The other is an invariance conclusion—the initial allocation is immaterial for whether a welfare-maximizing solution is reached. As just discussed, relaxing condition (iii) can undermine both conclusions, but, in a sense, it primarily undermines the invariance result. Efficiency, as judged by the Pareto criterion, will generally still be attained:

Theorem 2 (Modified Coase Theorem I) *Consider a bilateral contracting situation in which (i) the parties are rational with respect to their individual self-interests, but are not mean-spirited; and in which (ii) the parties can agree on any contract without incurring transaction costs. Then the allocation after contracting will be Pareto efficient regardless of the initial allocation.*

Proof: We need to show that the parties will never settle on an allocation, $\mathbf{a}$, that is Pareto dominated (note, now, an allocation may include the transferable good). Consider such an allocation. By definition, there exists at least one other allocation, $\mathbf{a}^*$, such that

$$u_i(\mathbf{a}^*) \geq u_i(\mathbf{a}) \tag{1}$$

[9] Willig (1976) provides justifications for assuming quasi-linear utility that are applicable in many contexts.

for both i and such that $u_i(\mathbf{a}^*) > u_i(\mathbf{a})$ for at least one i. Case 1: The inequality in expression (1) is strict for both i. Then $\mathbf{a}$ will not be implemented—both parties act rationally and they can costlessly contract, so they won't settle for $\mathbf{a}$. Case 2: (1) is an equality for one i. Then that i will refuse to implement $\mathbf{a}^*$ over $\mathbf{a}$ only if he is mean-spirited, which we have assumed he isn't. Hence, a Pareto-dominated allocation will not be implemented. ■

The modicum of altruism in the modified Coase Theorem—that the parties not be mean spirited—is unnecessary if we assume a strict-tradeoff condition:

Condition 1 (Strict Tradeoff) *Let $\mathcal{A}$ be the set of feasible allocations and $\mathbf{a}_1$ and $\mathbf{a}_0$ be two elements of $\mathcal{A}$. Then, if $u_1(\mathbf{a}_1) \geq u_1(\mathbf{a}_0)$ and $u_2(\mathbf{a}_1) \geq u_2(\mathbf{a}_0)$ with at least one inequality holding strictly, there exists an $\hat{\mathbf{a}} \in \mathcal{A}$ such that $u_1(\hat{\mathbf{a}}) > u_1(\mathbf{a}_0)$ and $u_2(\hat{\mathbf{a}}) > u_2(\mathbf{a}_0)$.*

The strict-tradeoff condition entails that if $\mathbf{a}$ is a Pareto-dominated allocation, then there exists a feasible allocation that strictly Pareto dominates $\mathbf{a}$; in which case, we need only Case 1 of the proof of the modified Coase Theorem and we can dispense with the case that relied on no mean-spirited behavior:

Theorem 3 (Modified Coase Theorem II) *Consider a bilateral contracting situation in which (i) the parties are rational with respect to their individual self-interests; in which (ii) the parties can agree on any contract without incurring transaction costs; and in which (iii) the set of feasible allocations satisfies the strict-tradeoff condition. Then the allocation after contracting will be Pareto efficient regardless of the initial allocation.*

With respect to freedom of contract, we have

Corollary 2 *Under the assumptions of the modified Coase Theorems, interference or restrictions on the contract the parties sign cannot increase the Pareto efficiency of the contracted-for outcome; that is, there should be freedom of contract if the only welfare issue is the efficiency of the outcome achieved by the contract from the perspective of the parties to the contract.*

Clearly, all the Coase theorems rely on the rationality assumption. We explore the consequences of relaxing this assumption later (see §2.3.4 and §4.2.1).

The no-transactions-cost assumption is also important. In the proofs, it serves to guarantee that the parties will not agree to a non-optimal contract because they can, without cost or impediment, choose an optimal contract instead. Relaxing this assumption would, thus, seem to have the potential for undermining the conclusions of these theorems—a point made by a number of authors (see, *e.g.,* Farrell, 1987b).

As Farrell notes, the economic literature on bargaining is generally sanguine about the prospects of an efficient outcome when the parties bargain under symmetric information. The literature is much more pessimistic, however, when

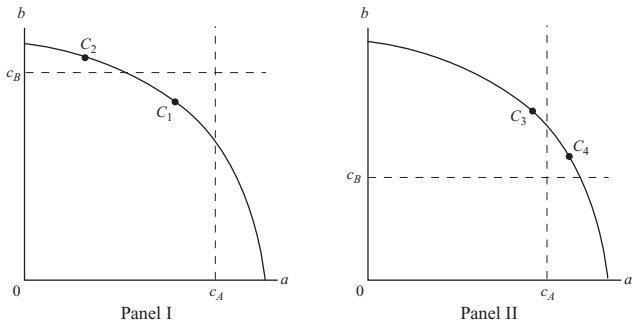


Figure 2: Consequence of transactions costs on bargaining outcomes under symmetric information.

they bargain under asymmetric information.[10] We discuss the consequences of bargaining under asymmetric information in §2.3.2. The rest of this section considers what can still potentially go wrong under symmetric information and concludes with Coase-theorem-like propositions that account for the potential of costly bargaining.

The potential consequences of transactions costs with symmetric information are illustrated in Figure 2. Suppose two parties, A and B, wish to enter into a contract. To do so, however, they must expend costs c_A and c_B, respectively. Once those costs are sunk (*e.g.,* lawyers are retained), the parties bargain over the contract terms. Let the payoffs, in money, be a and b, respectively. Because bargaining is conducted under symmetric information, theory predicts that the outcome will generally be efficient (see *e.g.,* Farrell, 1987b, or Sutton, 1986);[11] that is, the contract will be on the Pareto frontier, which is shown in the two panels of Figure 2 as a solid curve.

The first problem, illustrated in Panel I, is that c_A and c_B are so large that one or both sides prefer not to enter into negotiations. For instance, if bargaining would result in contract C_1, then neither side would be willing to negotiate; at C_1, we have $a < c_A$ and $b < c_B$. Moreover, while there are contracts that, if adopted, would induce one side to participate (*e.g.,* C_2, for which $b > c_B$), there

[10] See the survey on bargaining by Sutton (1986).

[11] Although efficiency is expected, it should be noted that one can construct perverse bargaining games that don't achieve efficiency even under symmetric information; see Hermalin and Katz (1993).

is no contract that would induce both to participate.

The problem illustrated in Panel I can serve to justify *default* contracts. That is, if the law stipulated that, absent a contract, the implicit contract was C_1, then that would clearly be an improvement. Observe that this improvement stems from two factors. First, the parties get to the Pareto frontier when they otherwise wouldn't. Second, they avoid the expenditures c_A and c_B. Note, however, that this analysis does *not* rely on C_1 being *mandatory*; this is not an argument for mandatory contracts.

A related problem, and one that *could* justify mandatory contracts, is illustrated in Panel II. Suppose that bargaining would lead the parties to contract C_3. Because $a < c_A$ at C_3, party A would refuse to enter into negotiations. This is undesirable, especially as there exist contracts, such as C_4, that, were they the outcome of bargaining, would induce both parties to negotiate. Now there is scope for limitations on contracts. If terms were limited, so that the only contracts that could be considered were on the arc segment between the dotted lines (*i.e.,* the segment containing C_4), then both A and B would be willing to enter into negotiations and they would get an outcome on the Pareto frontier.

Of course, one may wonder how the law would know that C_1 is a good default contract or to limit contracts to those in the neighborhood of C_4. If the law incurs costs arriving at default or mandatory contract terms, or lacks the information to design optimal contract, then it could be better to leave things in private hands. In other words, while it is true that restrictions on private contracts can possibly enhance efficiency when the private parties incur transactions costs, one must assess that observation in light of real-life limitations on what the legal system can do and the cost at which it can do it.

Before leaving the issue of transactions costs, it is worth considering how far we can go in relaxing the no-transactions-cost assumption and still establish a Coase-like case for freedom of contract. Our objective is to have a precise statement for a result of the form *if bargaining leads the parties to a second-best efficient*[12] *contract and does so in a way that minimizes bargaining costs, then the legal system can do no better than to leave the private parties' choice of contract unrestricted.*

To establish such a result, it is helpful to switch from working with allocations to working directly with contracts. Let C denote an arbitrary contract. Contract terms are assumed not only to fix the allocation of real goods (possibly contingently), but also the allocation of transfers (possibly contingently) between the parties. Let $U_i(C)$ denote party i's utility (possibly expected) should the parties agree to contract C.

We can restate the strict tradeoff condition as

[12]Second-best efficiency refers to the optimal outcome taking into account the informational constraints faced by the parties. For instance, in the standard hidden-action agency problem, the first-best outcome would entail the agent expending some ideal level of effort for a flat wage. But that outcome is typically infeasible once the constraint that the agent's action is hidden from the principal is taken into account; the second-best solution in such cases typically requires the agent to bear risk, which is inefficient from a first-best perspective.

Condition 2 (Strict Tradeoff) *Let $\mathcal{C}$ be the set of feasible contracts and consider any two contracts C_0 and C_1 in $\mathcal{C}$. If $U_1(C_1) \geq U_1(C_0)$ and $U_2(C_1) \geq U_2(C_0)$, with at least one inequality holding strictly, then there exists a contract $C_2 \in \mathcal{C}$ such that $U_1(C_2) > U_1(C_0)$ and $U_2(C_2) > U_2(C_0)$.*

Given the Strict Tradeoff condition, the following Coase-like theorem can be established:

Proposition 1 (Hermalin and Katz (1993)) *Suppose the two parties to a contract are symmetrically informed prior to and during bargaining and that bargaining consists of alternating offers. Assume the costs of delay in achieving an agreement are due to discounting. Assume that the set of possible contracts, $\mathcal{C}$, is invariant over rounds of bargaining and that the set $\{\big(U_1(C), U_2(C)\big) | C \in \mathcal{C}\}$ is convex and compact. Assume that there is at least one $C \in \mathcal{C}$ that each party strictly prefers to no agreement (the status quo). Finally assume either the Strict Tradeoff condition is satisfied or both parties are risk neutral and non-contingent (lump-sum) transfers in any amount are feasible. Then there is an essentially unique subgame-perfect equilibrium in which bargaining ends in the first round with agreement on a Pareto-efficient contract.*

The qualifier "essentially unique" captures the fact that there could be more than one contract that yields the unique equilibrium utility levels.

Proof: Follows from Lemma A2 and Proposition 4 of Hermalin and Katz (1993). ■

Because bargaining ends in the first round, there are no transactions costs on the equilibrium path. Hence, there is no possibility of increasing efficiency by reducing transactions costs. Moreover, because the resulting contract is efficient, there is no scope for increasing efficiency with respect to the contract chosen. In sum, under the assumptions of Proposition 1, there is no gain to be had from restricting private contracts; that is, Proposition 1 makes a case for freedom of contract.

On the other hand, Proposition 1 relies on a number of assumptions. Fortunately, some of these can be relaxed if one is willing to impose refinements on the subgame-perfect solution concept as applied to bargaining games. Specifically, we wish to rule out the possibility that the parties fear proposing efficient contracts because the equilibrium specifies a continuation game following the proposal of an efficient contract by party i that is unfavorable to party i (see Hermalin and Katz, 1993, for an example of such a perverse game). To that end, consider the following equilibrium refinements:

Condition 3 (Monotone Acceptance Condition) *Suppose that, at some date (round) t, a party would accept an offer of contract C_0. Suppose that another contract, C_1, would yield that party a higher expected utility level. Then, he or she will also accept contract C_1 at date t.*

Condition 4 (Stolen Thunder Condition) *Suppose that the equilibrium entails one party's making an offer of contract C at date (round) t on the equilibrium path for some $t > 1$. Then he or she would accept an offer of C by the other party in round $t - 1$.*

As Hermalin and Katz (1993) discuss, both refinements seem reasonable for bargaining games of complete information.

Given these refinements one can establish a stronger result:

Proposition 2 (Hermalin and Katz (1993)) *Suppose the two parties to a contract are symmetrically informed prior to and during bargaining and that bargaining consists of alternating offers. Assume the costs of delay in achieving an agreement are due to discounting or from per-round fixed costs. If the Strict Tradeoff, Monotone Acceptance, and Stolen Thunder conditions are satisfied, and there is at least one contract that each party strictly prefers to no agreement (the status quo), then bargaining ends with their agreeing to a Pareto-efficient contract in the first round of bargaining.*

Proof: This is Proposition 5 of Hermalin and Katz (1993). ■

Propositions 1 and 2 make the case that, when society is solely concerned with the wellbeing of the parties to the contract and those parties are symmetrically informed at the time of contracting, there is no reason to believe that restricting the parties' freedom of contract will improve efficiency.

2.3 The economic case against freedom of contract

So far, we have focused on the economic case *for* freedom of contract. Now we review the case against. Our discussion of the Coase Theorem suggests two potential grounds on which to argue against (complete) freedom of contract: (i) actors who are not party to a contract (third parties) are affected by externalities resulting from the contract; and (ii) problems in negotiating a contract prevent the parties from writing the optimal contract.

2.3.1 Third-party externalities

As we saw in §2.2.2, the efficiency of markets and private contracting is contingent on there being no third-party externalities. For instance, the market equilibrium with a competitive, but heavily polluting, industry does not maximize welfare—the supply of the good in question is determined by the private costs incurred by the manufacturers rather than the social costs that account for both those private costs and the harm the pollution imposes on society. Because social costs are greater than private costs, more than the welfare-maximizing quantity gets sold.

The inefficiency of the market when externalities are present can justify restrictions on private contracts. For instance, to deal with negative externalities, society does better by restricting the freedom of buyer and seller to set price; there exists a price floor above the market-equilibrium price that forces trade to occur at the welfare-maximizing level.[13]

A strong believer in the Coase Theorem might object to this conclusion, arguing that polluters and their victims could contract to set the pollution level optimally. While that might be plausible in the context of a single polluter and a single victim (*e.g.,* noise pollution issues between neighbors), most situations of interest involve multiple polluters and millions of victims. It is difficult to imagine that significant expenditures of time and effort aren't required for a multitude of parties to reach an agreement on the terms of a contract. Moreover, as Farrell (1987b) notes, the unknown intensity of the parties' preferences typically means that any such bargaining would occur under asymmetric information. When such real-life transactions costs are accounted for, restrictions on contracts could be the more efficient means of solving the externality problem.

The transactions-cost issue is worse when the victims of the externality are not even known at the time parties enter into a contract. Aghion and Bolton (1987) nicely illustrate the problem: A monopolist (*e.g.,* a manufacturer), concerned about entry into its market, signs long-term exclusive-dealing contracts with buyers (*e.g.,* retailers). An entrant enters only if it is more efficient (has a lower marginal cost) than the incumbent monopolist. Whether such an entrant will exist is unknown *ex ante* by the incumbent monopolist and any given buyer;[14] but both know the distribution of the potential entrant's marginal cost. Because of the potential for entry, the buyer can expect to earn more surplus in the future than it currently does; as noted by Bork (1978) and R. Posner (1976), this will make the buyer reluctant to enter into an exclusive dealing contract with the seller. However, some of the surplus generated by the entry of a more efficient entrant is captured by the entrant itself. Hence, if the buyer and incumbent seller collude—that is, enter into an exclusive-dealing contract with a liquidated-damages provision that the buyer must pay the incumbent seller if it switches to the entrant—then buyer and incumbent seller can capture some of the entrant's surplus. The entrant must lower its price *vis-à-vis* the price it would have charged absent any liquidated damages provision by the amount of the liquidated damages to induce the buyer to switch to it. If the entrant lowers its price, then surplus is being transferred from it to the buyer-seller combination. The problem with such exclusive-dealing contracts is that the buyer-seller combination is like a non-discriminating monopolist,[15] it sets the liquidated damages provision too high—in the combination's desire to capture more surplus from the most efficient entrants, it deters entry from those that

[13] A price floor is not the only policy tool that could improve efficiency relative to the unfettered market. Other possibilities are an excise tax on the good, a pollution tax, or permitting the manufacturers some ability to cartelize their industry.

[14] Later in their article, Aghion and Bolton relax this assumption and assume the incumbent monopolist has superior information; the implications of that assumption will be addressed later in §2.3.2.

[15] For a discussion of the welfare issues connected to a non-discriminating monopolist, see §2.3.3 *infra*.

are only moderately more efficient. Consequently, prohibiting exclusive-dealing contracts increases expected welfare by increasing the probability of entry by a more efficient producer.

Observe that it is difficult to invoke the Coase Theorem in response to the Aghion and Bolton model. Because the entrant is unavailable at the time the incumbent seller and buyer contract, there is no possibility of them signing a three-way contract that achieves efficiency.[16]

Following Rasmusen et al. (1991), exclusive dealing can illustrate another externality problem. Consider an incumbent monopolist who sells to N buyers. Normalize the surplus that each buyer enjoys under monopoly pricing to zero. Assume that, if there were entry, each buyer would enjoy surplus $s > 0$. Assume that entry is feasible only if an entrant can attract at least $\hat{N}$ buyers, where $1 < \hat{N} < N$. Observe, therefore, that if the incumbent can lock up at least $N - \hat{N} + 1 \equiv N^*$ buyers through exclusive-dealing contracts, the incumbent blocks entry. Consider a point in time prior to the arrival of an entrant; and consider the following offer made simultaneously by the incumbent to each of the N buyers: In exchange for signing an exclusive-dealing, the incumbent provides the buyer surplus $\varepsilon > 0$ (*e.g.,* the incumbent cuts its price by a small amount). The buyers respond independently and simultaneously to the incumbent. Take ε to be small enough that $N\varepsilon$ is smaller than the amount the incumbent would stand to lose should entry occur; that is, the incumbent does better paying out $N\varepsilon$ and keeping its monopoly than not paying that amount and facing competition. Note this will often entail $\varepsilon < s$.

Proposition 3 *There is a Nash equilibrium in which all N buyers sign an exclusive dealing with the incumbent.*

Proof: If a given buyer believes that the other $N-1$ buyers will sign, then that buyer believes entry has been blocked (recall $\hat{N} > 1$). Hence, that buyer expects to get 0 if she doesn't sign and ε if she does. Because $\varepsilon > 0$, it is, thus, a best response to sign. ■

The equilibrium of Proposition 3 is undesirable insofar as social welfare is reduced by the deadweight loss resulting from the preservation of monopoly pricing. Limitations on freedom of contract (*i.e.,* a prohibition on exclusive-dealing contracts) would be welfare enhancing.

As, however, Rasmusen et al. note, the equilibrium of Proposition 3 is not unique if $\varepsilon < s$. Another Nash equilibrium is for all buyers to refuse the contract—if no other buyer will sign, then signing would mean forgoing s in exchange for ε. Moreover, as Segal and Whinston (2000) point out, if one allows the buyers to form "coalitions," then the only Nash equilibrium will be the one in which all buyers refuse the contract.[17]

[16]There is also a further problem insofar as the entrant's cost is its private information, so there is an asymmetry of information problem that would impede efficient contracting even if the entrant were known *ex ante*. We elaborate on this point in §2.3.2.

If you, a buyer, sign and

1. if fewer than N^* buyers (including you) sign in total, then you will receive ε, but you will be *released* from the exclusive-dealing provision (*i.e.,* you can buy from the incumbent or the entrant); alternatively,
2. if exactly N^* buyers (including you) sign in total, then you receive $s + \varepsilon$ and you must buy from the incumbent; alternatively,
3. if more than N^* buyers (including you) sign in total, then you receive ε and you must buy from the incumbent.

Figure 3: A Clever Contract

Unfortunately, as Segal and Whinston go on to show, there is still scope for entry-deterring exclusive-dealing contracts. For instance, it is possible that N^*s is smaller than the amount the incumbent would stand to lose should entry occur. Hence, there exists an $\eta > 0$ such that $N^*(s + \eta)$ is smaller than the amount that entry would cost the incumbent. Then the incumbent could offer just N^* buyers an exclusive dealing contract in which each buyer received $s + \eta$. Because $s + \eta$ is greater than what a buyer receives even if entry occurs, it is a dominant strategy for each of these N^* buyers to accept this contract.

Indeed, building on ideas in Dal Bó (2003), it is possible for the incumbent to induce all the buyers to sign an exclusive-dealing contract at a cost to itself that is arbitrarily close to zero. Consider the contract in Figure 3.

Proposition 4 *If the incumbent offers the Figure 3 contract to all buyers, then it is a dominant strategy for each buyer to accept the contract. Hence, the unique Nash equilibrium is for all buyers to accept the contract, which entails the incumbent's paying a total of $N\varepsilon$ to block entry.*

Proof: Consider a given buyer and let n be the number of other buyers it expects to sign (so $N - n - 1$ is the number of other buyers it expects not to sign). There are three cases to consider:

1. $n < N^* - 1$. If the given buyer signs, she will be released from the exclusive dealing and her total surplus will be $s + \varepsilon$. If she doesn't sign, it will be just s. Hence, she should sign.

[17]The word coalition in this context has a specific game-theoretic meaning; roughly a coalition in this context refers to a self-enforcing agreement. That is, here, an agreement to reject the incumbent's offer would be self-enforcing because it is a Nash equilibrium for all buyers to reject. For the situations below in which the incumbent offers a contract to only N^* buyers or the incumbent uses a "clever contract," then an agreement among the buyers not to accept would *not* be self-enforcing—because it is a dominant strategy for some or all buyers to accept, they would not honor their agreement with their fellow buyers; that is, those contracts are robust to the formation of "coalitions." We elaborate on this point below.

2. $n = N^* - 1$ (*i.e.,* the given buyer is pivotal). If the given buyer signs, she will be obligated to buy from the incumbent, but she will be paid $s + \varepsilon$ in surplus. If she doesn't sign, then entry will occur, but her surplus will be just s. Hence, she should sign.

3. $n \geq N^*$. If the given buyer signs, she will be obligated to buy from the incumbent, but she will be paid ε in surplus. If she doesn't sign, then, because entry is blocked, she will get no surplus. Hence, she should sign.

Because, as shown, signing is best regardless of what the buyer thinks n is, signing is a dominant strategy. Because the given buyer was arbitrary, this holds for all buyers; that is, all buyers will sign. Because more than N^* buyers sign, the incumbent blocks entry, and does so at a cost of only $N\varepsilon$. ■

Although we have presented the inefficiencies illustrated by Propositions 3 and 4 in the context of exclusive dealing, they are, in fact, examples of a broader phenomenon. Segal (1999b) considers the general issue of a single contract proposer, P, who can enter into a number of bilateral contracts with other actors, $A_1, \ldots, A_N$, and derives conditions under which a set of unconstrained bilateral contracts will fail to maximize welfare due to externalities.[18] Other contexts include vertical relations (*e.g.,* exclusion of retailers or resale price maintenance), takeover battles (P is a raider and $A_1, \ldots, A_N$ incumbent shareholders), debt workouts (P is "equity," which offers a debt-equity swap to the creditors, the As), and network externalities (P sells a network good and the As purchase it).

Again, a strong believer in the Coase Theorem might object to these conclusions on two grounds. First, a single grand contract among all the participants would achieve efficiency if there is no asymmetry of information. Second, a *binding* agreement among the N actors (*e.g.,* the retailers) in advance of bilateral contracting with P (*e.g.,* the incumbent monopolist) would ameliorate, if not eliminate, the problem. There are, however, a number of counter-objections:

- If N is large, then the transactions costs are likely to be so large as to make contract restrictions more efficient.
- Some of the N actors could be unknown or not yet exist (*i.e.,* similar to the problem in Aghion and Bolton).
- Contracting among the As could generate other concerns. For instance, there could be legitimate antitrust concerns if all the retailers of a good or in an area were allowed to write a contract among themselves.

2.3.2 Asymmetric information

As has been known since Akerlof's (1970) seminal work, asymmetric information between parties can result in market distortions. A number of authors (Aghion and Bolton, 1987; Aghion and Hermalin, 1990; Johnston, 1990; Spier, 1992; Hermalin, 2002, among others) have applied this idea to the issue of contract design; showing that asymmetric information between the parties at the time a contract is negotiated can lead to distortions in the resulting contract *vis-à-vis* the contract that would have been negotiated under symmetric information. Unless the equilibrium under the symmetric-information contract is, itself, second best, such distortions must imply a loss of welfare *vis-à-vis* the symmetric-information benchmark.

Whenever the parties negotiate imperfect contracts, the question arises whether there is scope for the legal system to improve matters, either by restricting the set of possible contracts *ex ante* or through appropriate court action *ex post*. Aghion and Hermalin (1990) explore the former possibility in the context of a signaling game.[19,20] Their analysis can be motivated as follows. A well-known restriction on debt contracts is that the contract cannot contain an enforceable waiver by the debtor to her right to declare bankruptcy. Indeed, many states in the US impose even further protections, allowing a bankrupt debtor to keep certain assets (*e.g.,* a car or house) under certain circumstances. Can such restrictions enhance efficiency or total welfare?

To be concrete, consider an entrepreneur who needs to raise capital for a project. She knows how likely it is that her project will succeed; that is, whether it is a good project, which has a probability of failing of F_g or a bad project, which has a probability of failing of $F_b > F_g$. Using a short-hand common to information economics, call the entrepreneur with a good (alt. bad) project the good-type (alt. bad-type) entrepreneur. While the entrepreneur knows her type, a potential investor does not. His knowledge is limited to knowing that there is a probability $\theta \in (0, 1)$ that the project is good. Assume that if a project fails, it is impossible for the entrepreneur to repay the investor fully.

[18] For the latest on the general theory of bilateral contracting with externalities, see Segal and Whinston (2003).

[19] Signaling games, first studied by Spence (1973), are games of asymmetric information in which the better informed party takes actions that have the potential to convey—"signal"—her information to the less well informed party. The classic example (Spence) is a worker who signals information about her ability to potential employers through the amount of education she acquires. An equilibrium of a signaling game is called *separating* if the equilibrium actions of the informed player vary with her information (*e.g.,* workers who know themselves to be more talented acquire more education than workers who know themselves to be less talented). A *pooling* equilibrium is one in which the equilibrium actions of the informed player do not vary with her information (*e.g.,* all workers get the same level of education).

[20] Section II.B.2 and Appendix C of Johnston (1990) also considers the implications of signaling on contract formation in the context of evaluating limited-liability rules. Unlike Aghion and Hermalin, Johnston is concerned with *default* rules rather than binding restrictions. However, as Ayres and Gertner (1992) argue, Johnston's emphasis on default rules undermines his arguments; Ayres and Gertner show that the choice of *default* rule is irrelevant in a world with costless contracting.

A recent paper by Anderlini et al. (2003) is another contribution to this literature. They focus on *ex post* actions by the courts, specifically whether the court should void certain contracts in some states of the world. Unlike Aghion and Hermalin, who focus on how restrictions can shift the equilibrium from an inefficient separating equilibrium to a more efficient pooling equilibrium, Anderlini et al. show how the expectation of the court's *ex post* actions creates the possibility of an efficient separating equilibrium when otherwise the equilibrium would be a less efficient pooling equilibrium.

Investing is, therefore, risky and the risk is greater investing in a bad project than in a good project. For this reason, the entrepreneur can get more generous terms from an investor, the more likely he thinks it is that her project will succeed. Consequently, the entrepreneur has an incentive to signal the investor that her project is good through the terms of the debt contract she offers the investor. Specifically, because the expected cost of a large payment to be paid if the project *fails* is greater for a bad-type entrepreneur than a good-type entrepreneur, a good-type entrepreneur can signal that she has a good project by promising a large payment to the investor should the project fail. The cost of signaling in this manner is that the entrepreneur exposes herself to considerable risk (*e.g.,* losing her house if the project fails).

Restricting the amount the entrepreneur can promise to repay in the case of failure can potentially generate Pareto superior outcomes. To see why, note that, because of the additional risk, an entrepreneur with a good project might prefer not to signal if silence were interpreted by the investor as meaning her project was "average" (*i.e.,* had a failure probability of $\theta F_g + (1-\theta)F_b$). The difficulty is that, under a reasonable solution concept for the game,[21] the investor will interpret silence as evidence that the project is *bad*; and given the choice between looking good (signaling) and looking bad (not signaling), an entrepreneur with a good project will prefer to look good. If, however, signaling is restricted (*e.g.,* bankruptcy laws limit what the entrepreneur can pay in the event of failure), then not signaling is no longer informative. The investor will, therefore, treat all entrepreneurs as if they have an average project. Both types of entrepreneur are better off—an entrepreneur with a bad project now looks average, while an entrepreneur with a good project avoids the additional risks imposed by costly signaling. Because the investor is always held to his reservation utility conditional on his equilibrium beliefs, he is no worse off. Thus, restricting the possible terms of the contract would be Pareto superior.

Aghion and Hermalin formalize this argument in the entrepreneur-investor context. They go on to suggest, but not model formally, that the idea of contract restrictions eliminating "wasteful" signaling is more general than the entrepreneur-investor example. For instance, it could justify limits on penalties for breach of contract: To signal that she is very likely to be able to deliver a product on time, a good-type supplier might offer "too high" a penalty to be paid were she to be late; where "too high" means that she would be happier with a lower promised penalty if only the buyer wouldn't interpret that lower penalty as indicating she was the bad-type supplier. Barring excessive penalties would prevent the buyer from making that interpretation, which in turn would lead to a contract that both good and bad-type suppliers preferred.

While Aghion and Hermalin prove that restrictions on contracts *can* be welfare enhancing in the context of signaling models, they do not establish that restrictions will always be welfare enhancing. For some set of parameters, restrictions enhance welfare; for others, they don't. Moreover, in the latter

[21] Specifically, a Perfect Bayesian Equilibrium (PBE) satisfying the Intuitive Criterion of Cho and Kreps (1987).

case, the imposition of binding restrictions will reduce welfare (see Figure 5b of Aghion and Hermalin and connected discussion). Intuitively, there exist parameter values such that the separating PBE under asymmetric information replicates the equilibrium that would hold under symmetric information. In those situations, given our earlier discussion, it is not surprising that restrictions can only reduce, not enhance, welfare. Furthermore, if the contracting situation is already problematic, the fact that the informed player must signal can improve matters: Aghion and Bolton (1987), for instance, point out that the introduction of asymmetric information in their model pushes down the average liquidated damages penalty, thereby increasing the likelihood of efficient entry.

The lack of a clear normative conclusion from Aghion and Hermalin admittedly limits the practical application of their results. Eric Posner (2003a) sees this as a fatal flaw, supporting his overall indictment of contract theory as a guide to the law of contracts. But, as Craswell (2003) notes, there is no reason to require economic analysis to reach "all or nothing" conclusions before the analysis is useful normatively.

A signaling game is a particular kind of game of asymmetric information; in the context of contract design, it corresponds to a situation in which the informed party (*e.g.,* the entrepreneur) makes a contract offer to the uninformed party (*e.g.,* the potential investor). But there are other possible contract-offer games. An obvious alternative is for the uninformed party to make a contract offer to the informed party (a game known as *screening*). Can restrictions on contracts improve the efficiency of the outcomes of screening games?

The answer, as demonstrated by Hermalin and Katz (1993), is effectively no. The argument is as follows, let $\mathcal{C}_U$ be the set of unrestricted contracts and let $\mathcal{C}_R$ be the subset of restricted contracts (*i.e.*, $\mathcal{C}_R \subset \mathcal{C}_U$). Let C_i^* be the contract offered by the uninformed party and accepted by the informed party in the game where the relevant contract space is $\mathcal{C}_i$.[22] Let $v_P(C)$ be the uninformed party's expected utility under contract C. Because the *un*informed party cannot signal information, changing the contract space cannot change the informed party's acceptance rule. Hence, by the nature of optimization,

$$v_P(C_U^*) \geq v_P(C_R^*)\,. \tag{2}$$

Clearly, if the inequality is strict, then restrictions on contracts cannot be Pareto improving. If expression (2) is an equality, then there is no reason for the uninformed party not to offer C_R^* if it Pareto dominates C_U^* and, hence, it is unclear why we shouldn't expect C_R^* to be offered even absent restrictions. Moreover, if the Strict Tradeoff condition (Condition 2) holds, then C_R^* cannot Pareto dominate C_U^* if expression (2) is an equality. To see this, were C_U^* dominated by C_R^*, then, by the Strict Tradeoff condition, there would exist

[22] As is typical in contract theory, we can, without loss of generality, add the "refusal contract" to the set $\mathcal{C}_R$ (and, thus, to $\mathcal{C}_U$); where the refusal contract stipulates the same payoffs to the parties as would result if the informed player refused the uninformed player's offer. In other words, there is no loss of generality in assuming acceptance of some contract in equilibrium.

a third contract $\hat{C} \in \mathcal{C}_U$ that the uninformed player strictly preferred to C_U^* and which the informed player would accept. But the existence of such a $\hat{C}$ contradicts the optimality of C_U^*. Therefore, C_R^* cannot Pareto dominate C_U^*. A robust conclusion, therefore, is

Proposition 5 *Restrictions on contracts cannot be Pareto improving in screening situations if either (i) the uninformed party is not mean spirited; or (ii) the contract space satisfies the Strict Tradeoff condition (Condition 2).*

Why do signaling and screening models yield different conclusions? Externalities offer an explanation. Although the informed player in a signaling game is a single entity, one can nonetheless view her as being two (*e.g.*, the bad-type entrepreneur or the good-type entrepreneur). The fact that a bad type is potentially willing to pretend to be a good type forces the good type to distort the contract she offers so that she won't be mistaken for the bad type (*i.e.,* select a contract, though not ideal for her, that the bad type would *not* be willing to offer). The problem is that there is no way for the bad type to internalize this externality that her potential mimicry imposes on the good type. When, however, it is the *un*informed party who makes the contract offer, he is in a position to internalize the costs and benefits of attempting to differentiate the different types of the informed player. In essence, signaling can impose an externality, while screening cannot.

We can also consider social welfare in screening models of contract bargaining. Obviously, if restrictions are Pareto improving they also enhance welfare. However, as we've just seen, it will generally be the case that restrictions will not be Pareto improving with screening. Hence, we limit attention to the case in which expression (2) is a strict inequality. The question of whether restrictions can enhance welfare then boils down to whether the informed party's (average) gain from the restrictions exceeds $v_P(C_U^*) - v_P(C_R^*) > 0$.

To see that restrictions can enhance welfare, consider the following example. The uninformed player is a seller and the informed player is a potential buyer. The buyer's private information is his knowledge of the benefit, b, he derives from a single unit of the good being sold by the seller. Assume that b is drawn prior to contracting from a uniform distribution on $[0, 1]$ and this fact is common knowledge. For convenience, assume the good holds no intrinsic value for the seller, so her cost is zero. While the buyer knows his benefit from purchase at the time of contracting, the seller knows only it was drawn from the uniform distribution. The unconstrained contract offered by the seller will be a contract to sell the good at a price of $1/2$, which yields total expected welfare of $3/8$ (of this $1/4$ is expected profit for the seller and $1/8$ is the expected surplus captured by the buyer).[23] Consider the restriction that $p \leq 0$. Within this constrained

[23] At price p, the buyer will buy if $b \geq p$. The probability $b \geq p$ is $1 - p$. The seller's expected profit from a price p is, thus, $p(1 - p)$, which is maximized by $p = 1/2$. If the unit is traded, welfare is just b; hence, expected welfare is $\int_{1/2}^{1} b db = 1/2 - 1/8 = 3/8$.

space, the seller will set a price of $p = 0$. All types of buyer buy, so expected welfare is $1/2$ (all of which is captured by the buyer).[24]

Observe that the welfare loss that arises in an unrestricted world occurs because of asymmetric information. If the seller knew the buyer's valuation, then the seller would set $p = b$. All types of buyer would buy, so expected welfare would be $1/2$ (all of which would be captured by the seller). Alternatively, if the buyer did not know his valuation at the time of contracting—so the parties are symmetrically informed insofar as they both know only that $b \sim \mathsf{U}[0, 1]$—then welfare would be maximized by a contract that set $p = 1/2$ and the buyer always bought; the seller gets $1/2$ for sure and the buyer's expected surplus is zero, so total welfare is $1/2$. Note this last result is consistent with the view that asymmetries of information that arise *after* contracting are *not* justifications for restrictions; a point made more generally by Hermalin and Katz (1993)—see Propositions 1 and 2 above.

In §2.2.1, we observed that the welfare criteria of Pareto efficiency and social welfare often coincide when transfers between the parties are feasible. It is worth, therefore, considering why that coincidence breaks down with asymmetric information. When there is asymmetric information, transfers are called upon to serve double duty. They continue to be a means of transferring surplus so that the welfare-maximizing allocation might be viewed as Pareto efficient by the parties. But, with asymmetric information, they are also a means of screening the different types. As our simple example illustrates, this second duty impedes transfers from doing the first optimally.

To summarize: In a signaling situation, restrictions on contracts can lead to Pareto superior outcomes and, thus, can increase total welfare. In contrast, in a screening situation, restrictions on contracts generally cannot be expected to generate Pareto improvements, although they can increase total welfare.

2.3.3 Market power

As discussed above, competitive markets can be expected to maximize welfare in the absence of externalities. When, however, one or more entities have *market power*, the market can no longer be expected to yield the social welfare-maximizing allocation. It is well known that a firm with market power (*i.e.,* that faces a downward sloping firm-specific demand) produces less than the welfare-maximizing quantity. So, at least in the standard static framework, market power reduces welfare.[25] Consequently, public policy should generally oppose contracts that promote market power over competition, such as cartel agreements. This logic extends to other contracts, such as exclusive-dealing con-

[24] Expected welfare is $\int_0^1 b db = 1/2$.

[25] In a dynamic framework, one may need to consider the incentive effects of monopoly profits for innovation; that is, in some contexts, without the monopoly profits that intellectual property protection (*e.g.,* patents) afford, the innovator would lack sufficient incentive to innovate. No innovation means zero units are traded, which yields even less welfare than the monopoly outcome.

tracts,[26] that firms might sign to maintain, establish, or extend market power.

It is important to recognize that the welfare loss that comes from market power (*i.e.,* non-discriminating or simple monopoly pricing) is an example of the adverse welfare consequences of asymmetric information in a screening model (see §2.3.2).[27] While buyers know their valuations for each unit, the monopolist does not. Hence, the monopolist sets her price both to affect the transfer of surplus from buyers to herself and to screen the buyers. As before, asking a single instrument to serve two roles leads to distortions in welfare.

The fact that the monopoly-pricing problem is a screening problem also implies that if the monopolist knew each buyer's valuation schedule for all the units he could conceivably wish to purchase, then it could achieve the welfare-maximizing allocation. That is, as is well known (see, *e.g.,* Tirole, 1988, §3.1), a perfect price-discriminating monopolist maximizes social welfare.[28] Hence, the welfare loss typically seen with monopoly stems not from market power alone, but from the combination of market power and asymmetric information.

Price discrimination in the real world is imperfect. Depending on the form of discrimination, the structure of demand, and other circumstances, allowing a monopolist to engage in price discrimination versus simple monopoly pricing can enhance or diminish welfare (see Chapter 3 of Tirole, 1988, for a review of the welfare consequences of imperfect price discrimination). Consequently, it is difficult to assess, at a general level, what policy should be towards the enforceability of contractual terms that facilitate price discrimination.

For example, it is not obvious from a welfare perspective whether airlines should be free to issue tickets that contain restrictions (*e.g.,* an obligation to stay a Saturday night before returning to one's point of origin).[29] Such matters need to be studied on a case-by-case basis.

Note too that market power is connected to bargaining power. Indeed, the screening problem associated with a non-discriminating monopolist can be interpreted as stemming, in part, from the seller having all the bargaining power,

[26] Recall the discussion of Aghion and Bolton (1987) and Rasmusen et al. (1991) in §2.3.1.

[27] It follows from the Coase Theorems (Theorems 1–3) or Propositions 1 & 2, as appropriate, that bargaining power should *generally* be irrelevant in a standard welfare analysis *absent* asymmetries of information.

This is not to say that there couldn't be other social reasons for concern about inequities in bargaining power, such as distributional concerns (consider, *e.g.,* Kennedy, 1982). As E. Posner (2003a) notes, courts have been known to declare contract unconscionable because of unequal bargaining power. See also §2.4.1 and §2.5.2 *infra*.

[28] Of course, because a perfect-discriminating monopolist captures 100% of the surplus, there could be distributional or equity grounds for this outcome not to be favored.

[29] Saturday night restrictions are a form of second-degree price discrimination whereby an airline can screen business travelers (those with a high value of flying and a high cost of staying over Saturday) from non-business travelers (those with lower values of flying and lower costs of staying over Saturday). If banning such restrictions caused the airlines to price so that far fewer non-business travelers flew, then the ban would almost surely be welfare reducing. If, however, the ban led the airlines to price so as to keep non-business travelers flying, then eliminating the distortionary effects of the Saturday-night restriction would almost surely be welfare enhancing.

so that she gets to make a take-it-or-leave-it offer to buyers. The welfare loss from monopoly would disappear if it were the buyers who could make take-it-or-leave-it offers (assuming they knew the seller's marginal-cost schedule).

2.3.4 Capacity and bounded rationality

The parties in traditional law & economics analyses are presumed to be sophisticated and possess the requisite capacity. Consequently, any contract into which they voluntarily enter is, in rational expectation, superior for them to their no-contract (*status quo*) position. That is, each party correctly estimates that their expected utility from contracting exceeds that from not.

Observe that the rationality being assumed has two components. First, neither party would enter into an agreement that he or she thought would make him or her worse off, in expectation, than not contracting. Second, each party is forming these expectations in an objectively correct manner. For instance, if you respond to some get-rich-quick spam email, you presumably expect to enrich yourself, but such expectations are not rational; that is, you are rational in the first sense, but not the second.

It is difficult to argue that people aren't rational in the first sense (putting aside certain pathologies such as compulsive self-destructive behavior), but there has long been some unease with the assumption that people are rational in the second sense (see, *e.g.,* Simon, 1972, or Rubinstein, 1998, for a discussion). More recently, a movement has arisen within law & economics generally (see, *e.g.,* Jolls et al., 1998; Korobkin and Ulen, 2000) that also questions the assumption of rationality in the second sense. This movement has been labeled "behavioral law and economics," and it resembles related work in economics in its emphasis on cognitive errors, framing effects, time-inconsistent patterns of discounting, and similar phenomena. Within the formal modeling of contracts, however, assumptions of *bounded rationality* have been largely limited to the issue of explaining incomplete contracts (or justifying assuming that contracts are incomplete). We take up this use of bounded rationality in §4.2.1 below.

While, to the best of our knowledge, capacity and sophistication have not been formally modeled in the context of contract theory, such issues have received attention by law & economics scholars in law reviews (see, *e.g.,* Eisenberg, 1995; Korobkin and Ulen, 2000; Bar-Gill, 2005).[30] Korobkin and Ulen, for instance, argue that mandated contractual terms can be justified when one side lacks sophistication. They take as an example the former practice of insurance companies to cover only one day of hospital stay for maternity and the legislative reaction that required these companies to cover longer stays. They argue that legislative action was necessary because "the possible permutations of coverages that could, in principle, be provided by a given health insurance policy are numerous and ... can overwhelm even sophisticated consumers ... [and their] agents ... who might be making purchasing decisions" (Korobkin and

[30] Although, as we discuss *infra* some formal models (*e.g.,* Katz, 1990b; Rasmusen, 2001) touch on related issues.

Ulen, p. 1082).

While it is difficult to argue against the proposition that the majority of consumers do not fully understand the provisions of their insurance contracts,[31] the example of maternity benefits seems a poor one from which to argue for intervention on the basis of inattention. The legislative reaction to this issue, and the strong public outcry that prompted it, suggest that the limited coverage insurers had previously been offering was not due to any consumer inattention. A more likely explanation was insufficient willingness to pay.[32] Put a bit differently, given the apparent awareness of the market to how much maternity stay was being covered, one would have expected the "unregulated market" to provide the desired level of maternity benefits if maternity benefits had been important to those who actually decided about policies.

A more plausible variant of the Korobkin and Ulen idea is developed by Eisenberg (1995). Although Eisenberg presents his argument verbally rather than mathematically, it is helpful to add some degree of formalism.[33] Suppose there are two possible future states, a rare one, which occurs with probability r; and a more likely one, which occurs with probability $1-r$. Eisenberg is interested in the case where r is small, but not zero. Consider contracting between two parties, P and A. For convenience, assume that the bargaining between them always results in a contract that yields monetary benefits B_i (gross of direct transfers) to party i in the more expected state (*i.e.,* the state that occurs with probability $1-r$). Suppose that the contract can contain a stipulation, $s \in \mathcal{S}$, that governs what happens in the rare state. Let $b_i(s)$ denote party i's monetary benefit gross of direct transfers in the rare state under stipulation s. Assume that the stipulation that maximizes P's benefit in the rare state, $\hat{s}$, is not the stipulation that would maximize the sum of the parties' benefits. Let

[31] See, for example, Eisenberg's (1995) discussion (p. 242) of *Gerhardt v. Continental Insurance Cos.* (48 N.J. 291, 295 A.2d 328), in which he quotes the justices of the New Jersey Supreme Court expressing their difficulties in understanding the terms of the insurance contract in question.

[32] Korobkin and Ulen cite statistics (their footnote 111) that 90% of insured Americans get their health insurance through their employer. This fact, however, is potentially problematic for their argument. The employers could be quite sophisticated and are simply shopping for a bargain in insurance. That is, the problem is due to agency (the employers obtaining insurance on their workers behalf) rather than bounded rationality. Korobkin and Ulen might argue that workers suffer the same cognitive limitations bargaining with their employers; but, depending on worker-employer bargaining, the relative political power of workers versus employers and insurers, and the incidence of the cost of increased medical coverage, it is nevertheless possible that everything can be explained within the rational-actor paradigm.

[33] Katz (1990b) and Rasmusen (2001) are two articles that also model "reading costs" formally.

that stipulation be s^*. Mathematically,

$$\begin{aligned} b_P(\hat{s}) &\geq b_P(s) \text{ for all } s \in \mathcal{S}\,, \\ b_P(\hat{s}) &> b_P(s^*)\,, \\ b_A(s^*) + b_P(s^*) &\geq b_A(s) + b_P(s) \text{ for all } s \in \mathcal{S}\,, \text{ and} \\ b_A(s^*) + b_P(s^*) &> b_A(\hat{s}) + b_P(\hat{s})\,. \end{aligned} \tag{3}$$

Observe that there must be $s \in \mathcal{S}$, including s^*, that A prefers to $\hat{s}$. Define $\Delta_i = b_i(s^*) - b_i(\hat{s})$. By construction,

$$\Delta_A > \Delta_A + \Delta_P > 0 > \Delta_P\,.$$

Eisenberg is interested, in part, in situations in which P proposes the stipulation for the rare state and it costs A some amount, k, to evaluate or understand the stipulation that P has proposed. One can view k as the cost of consulting an attorney or the cost of time spent trying to understand the small print. If A is particularly unsophisticated, one might even set $k = \infty$ to reflect the idea that A is simply incapable of understanding what P has proposed. Suppose that if A does understand the stipulation, then P and A would bargain to some term that A found acceptable. In keeping with our earlier analysis of bargaining under symmetric information, we may as well assume that they would then agree on s^*. A critical assumption, however, is that

$$k > r\Delta_A\,; \tag{4}$$

in other words, the expected gain that A stands to reap from understanding the stipulation for what should happen in the rare state is smaller than the cost of obtaining that understanding (we will see in a moment that, when A doesn't understand the stipulation in the rare state, P should propose $\hat{s}$). Observe that expression (4) holds if k is large or if r or Δ_A are small; Eisenberg can be read as being primarily concerned about the situation in which r is small—we will return to this point later.

To close the model, suppose that the bargaining process is such that P can be seen as proposing a contract subject to A's acceptance, where A accepts only if his expected net benefit at least meets a reservation level β_A; that is,

$$(1-r)B_A + r\tilde{b}_A - p \geq \beta_A\,, \tag{5}$$

where $\tilde{b}_A$ is the benefit A *expects* in the rare state and p is a transfer to P (if $p < 0$, then the transfer is from P to A). P's utility is

$$(1-r)B_P + r\tilde{b}_P + p\,,$$

where $\tilde{b}_P$ is what she gets in the rare state.

Suppose that A will not expend the k necessary to understand the stipulation for the rare state; then it is irrelevant for A's acceptance rule (5) what stipulation P *actually* makes because A won't understand it and, thus, his expectation,

$\tilde{b}_A$, cannot depend on it. Given this, it is clearly a best response for P to choose $\hat{s}$.[34] In equilibrium, A must form his expectations correctly,[35] so $\tilde{b}_A = b_A(\hat{s})$. Observe, given (4), it is indeed a best response for A not to expend k on understanding even if A anticipates P is proposing $\hat{s}$. In equilibrium, total welfare will, thus, be

$$\widehat{W} \equiv (1-r)(B_A + B_P) + r\big(b_A(\hat{s}) + b_P(\hat{s})\big)\,.$$

From expression (3), it follows that total welfare is *not* being maximized.

If we suppose that P has sufficient bargaining power so that expression (5) is binding, then P's equilibrium utility is

$$(1-r)(B_A + B_P) + r\big(b_A(\hat{s}) + b_P(\hat{s})\big) - \beta_A\,,$$

from which we see that it is P who bears the full cost of this loss in welfare. Normally when a party bears the full cost, that party would take steps to eliminate the efficiency loss. However, here, P can't—there is no credible way for her to commit to A that she has stipulated s^*; that is, any promise of s^* is cheap talk because P knows A will not understand if P substitutes the stipulation that favors her.

There are some ironies in this result. First, while Korobkin and Ulen and Eisenberg appear to suggest that it will be the *un*sophisticated party who suffers in these situations, the truth could well be that it is the sophisticated party who suffers. Second, even though, as noted by Schwartz and Wilde (1979), the sophisticated party has a motive to internalize the value of the superior contract term, there is no way for her to do so.[36]

Conversely, suppose that, due to competition among the Ps, it is the As who capture the surplus; that is,

$$(1-r)B_P + rb_P(s) + p = K_P\,,$$

where K_P is P's reservation utility (cost). Observe this expression can be rearranged as

$$p(s) \equiv K_P - (1-r)B_P - rb_P(s)\,. \tag{6}$$

From expression (6), it follows that the price a P charges if the term is $\hat{s}$ is lower than if it is s^*; that is, $p(\hat{s}) < p(s^*)$. In this environment, we can assess

[34] One might wonder why P bothers to worry about what to stipulate if the state really is rare. One answer, suggested by Eisenberg, is that P deals with many As (*e.g.,* the contract is a form contract that P uses with many customers); hence, even though the expected gain to P from stipulating $\hat{s}$ over s^* is small for any one transaction, in aggregate it is large enough to induce P to invest in choosing the best stipulation for her.

[35] This imposition of rationality on A might seem at odds with the bounded rationality approach of Eisenberg. We will discuss this point later.

[36] In Katz (1990b), the sophisticated party can take actions costly to it that lower the unsophisticated party's cost of understanding the contract terms. On the other hand, if those costs are incurred on a per-transaction basis (*e.g.,* they represent the expense of a person explaining terms to a potential buyer), then such actions could well fail to be cost effective, especially if the terms in question pertain to a rare event.

Schwartz and Wilde's claim that if some of the As have a low enough cost of evaluating the stipulation for the rare state, then competition will lead the Ps to stipulate s^* rather than $\hat{s}$. Suppose there are two kinds of As, those with low evaluation costs, k_ℓ, and those with high evaluation costs, k_h. For the moment, set $k_\ell = 0$, so the ℓ-type As always evaluate. For the h-type As, k_h satisfies expression (4). Suppose the presence of the ℓ types led all Ps to stipulate s^*. Competition among the Ps would then require them each to charge $p(s^*)$ and earn zero profits. Consider a deviation from this candidate equilibrium in which a P stipulated $\hat{s}$. Although she would lose all business from ℓ types, she would, at least in the short term, keep the h types. Moreover, at a price of $p(s^*)$, she would realize an expected gain of $-r\Delta_P > 0$ from a contract with an h type. Hence, deviating from the candidate equilibrium—stipulating $\hat{s}$ rather than s^*—is profitable for that P, which means the candidate equilibrium isn't an equilibrium at all. In other words, there is no equilibrium in which all Ps stipulate s^* for sure.[37]

In fact, if we set $k_\ell > 0$, we can see a second reason why an s^*-only equilibrium cannot arise. Define $\tilde{\Delta}_A = b_A(s^*) - \tilde{b}_A$; that is, $\tilde{\Delta}_A$ is the expected gain for A that he expects from negotiating the contract from the expected stipulation in the rare state to s^*. Observe, even for an ℓ type, it is only worth evaluating the stipulation if

$$k_\ell \leq r\tilde{\Delta}_A\,.$$

Suppose all Ps were expected to stipulate s^*; then, $\tilde{\Delta}_A = 0$. Hence, no types of A would bother to evaluate the stipulation. But, as we've seen, if A won't evaluate, then P should always stipulate $\hat{s}$. The conclusion is, thus, that it is impossible for A to ever be certain that P has stipulated s^* unless A checks (we might dub this the "Eisenberg uncertainty principle").[38]

Although this model is somewhat rudimentary, it does demonstrate that the problem identified by Eisenberg and others is unlikely to be eliminated by private action. This suggests that intervention could be beneficial. Such intervention could be *ex ante*, *i.e.,* the law could disallow any stipulation but s^*; or the intervention could be *ex post*, *i.e.,* the courts could grant A damages should it be shown that the stipulated term was other than s^*.

As we observed earlier, a critical condition is (4), which makes it rational for A not to seek to understand the terms stipulated for the rare state. If condition (4) fails, as it will if the relevant terms (r or Δ_A) are large, then the argument for intervention collapses.[39] While either r or Δ_A small enough is sufficient for condition (4) to hold, the situation in which it holds because Δ_A is small

[37] What *is* the equilibrium in this game depends on what additional assumptions we make. If, for instance, we assumed that ℓ types could, in negotiations, change the contract from $\hat{s}$ to s^* in exchange for an increase in price from $p(\hat{s})$ to $p(s^*)$, then an equilibrium can be constructed in which all Ps offer contracts at $p(\hat{s})$, but end up negotiating different terms with the ℓ-type As.

[38] Katz's Lemma 2 makes the same point.

[39] Although, even when (4) fails, there could still be cause to be concerned due to the Eisenberg Uncertainty Principle.

could be relatively uninteresting—presumably A will need to take some action to enforce restrictions on s (*e.g.,* sue to have an illegal contract voided or to recover damages); but if Δ_A is small, then A's motive to do so could be small as well. That is, it could be difficult to implement intervention in cases in which Δ_A is small. The situations in which r is small are, as Eisenberg suggests, the more important ones. The *ex post* injustice could be large, but the motive of a single A to address it *ex ante* could be too small to make feasible a private solution.

Admittedly, we have imbued A in our formal model with rather more rationality than Eisenberg intended. Like others (*e.g.,* Jolls et al., 1998; Korobkin and Ulen, 2000; Bar-Gill, 2005), he observes that research in psychology and behavioral economics suggests that most people are poor at estimating probabilities correctly, especially probabilities of small events.[40] To the extent that they systematically underestimate the probability of a rare event (*e.g.,* likelihood of an insurance claim), they will use too small an r in deciding whether to invest in understanding what the contract stipulates for rare states. While the addition of such cognitive biases have the potential to enrich the analysis, our simple model demonstrates that they are not essential to obtain the main conclusion.

Katz (1990b), among others, notes there is a relation between the analysis in this subsection and Akerlof's lemons model. Although, here, we have a problem of moral hazard—the contract proposer *chooses* her "type"—whereas Akerlof is concerned with a problem of adverse selection—the informed party's type is *exogenously* determined—it is true in both that the uninformed party expects adverse terms of trade. A major difference, however, is that here all valuable exchanges take place, but just not under the optimal contract. In the lemons model, in contrast, buyer suspicion about quality can push the price below the level at which high-quality sellers are willing to sell, leading to inefficient exit of the high-quality sellers from the market. Admittedly, if the terms in question were important enough to the uninformed party, but yet not important enough to incur the reading costs, then it is theoretically possible that no trade could occur because of the uninformed's suspicions about the adverse nature of the contract. These issues have been discussed informally in the legal literature as well (see, for example, Kennedy, 1982; Craswell, 1993, 2000; and Eisenberg, 1995).

2.4 Other arguments for regulating private contracts

2.4.1 Distributional fairness

The Second Theorem of Welfare Economics holds that given convex production and utility functions, complete markets, and costless redistribution, any Pareto efficient allocation can be implemented as the general equilibrium of a competitive economy (see, *e.g.,* §6.4 of Debreu, 1959). In the real world, however, these

[40] See Rabin (1998) for a survey.

necessary assumptions are frequently violated, so that the same policy instruments must be used to promote both efficiency and distributional justice, with the two goals being traded off in a second-best fashion (Okun, 1975). Accordingly, much economic regulation has in practice some distributional component (R. Posner, 1971, and Polinsky, 1974).

In the field of contracts, however, as opposed to other bodies of law such as tort or property, it is relatively difficult to use private law rules as instruments of distribution. This is because contractual obligations are in general voluntarily undertaken, so that changes in legal rules will be accompanied by changes in the parties' reservation prices for exchange, tending to shift the incidence of a regulation to its intended beneficiary (*e.g.,* Buchanan, 1970). But restrictions on price will redistribute between marginal and inframarginal market actors, and in cases of market power, can redistribute between the two sides of the market (as can readily be seen from the example of a monopolist subject to a price ceiling set between the monopoly price and marginal cost). And restrictions on non-price terms can similarly redistribute between marginal and inframarginal actors if they differ in the value they attach to such terms (Spence, 1975; Craswell, 1991).

Similarly, policies that internalize externalities have distributional effects as well, and from a political economy viewpoint such effects—because they are larger—are probably more important than any efficiency gains in explaining the pattern of regulation. For example, in situations of adverse selection (the most important externality arising in the contractual setting), policies that force information revelation or limit entry or exit from a market can cause redistribution across informational types.

2.4.2 Liberty and autonomy

In the context of contract law, the liberal perspective tends to support arguments for contractual freedom, although on libertarian rather than economic grounds. As indicated in §1.2.3, however, liberal and economic theories of contract can diverge in cases of market failure or transaction costs. A canonical illustration is provided by the phenomenon of standard-form contracts. Standardized contracts are an inevitable by-product of a mass production economy, but many legal commentators (*e.g.,* Kessler, 1943; Rakoff, 1983) have regarded them with distrust on the grounds that most persons presented with standard forms quite reasonably do not bother to familiarize themselves with the specific contents, relying instead on the drafter's reputation and on the knowledge that other contracting parties regularly do business on like terms. Such commentators take the position that these facts vitiate the non-drafting party's consent and justify state regulation of fine-print terms, although other writers, more influenced by an economic or commercial perspective (*e.g.,* Llewellyn, 1960, pp. 362–371) are willing to read implied consent into the overall situation. This issue is discussed at greater length in §3 and §4 below, where we take up the problems of contract formation and interpretation.

2.4.3 Inalienability and commodification

Finally, both popular morality and legal institutions commonly limit transactions dealing with matters thought to be fundamental to citizenship or personal identity; common examples include prohibitions on slavery, sexual prostitution, and the transfer of political rights such as suffrage or military service. The entitlements subject to such restriction are often described in political terms as inalienable, a concept not easily incorporated into economic accounts of exchange. Sometimes, as in the case of sexuality or body parts, restrictions are imposed on market exchange but not on other transfers: a situation that Radin (1987) labels "market-inalienability."

In some cases, restrictions on alienability can be justified in terms of market failure such as asymmetric information (*e.g.,* the use of in-kind benefits to screen out those not in need, as in Blackorby and Donaldson, 1988) or externality (when Pigouvian taxes are administratively infeasible). But many such restrictions are better explained by the idea that exchange of the relevant entitlement injures some fundamental interest of the restricted agent not captured by his utility function, or some social interest that cannot be translated into material or pecuniary terms. In economics, this idea finds historical roots in the classical Marxian idea of commodification, or the change in the social meaning of a good that occurs when it becomes the subject of economic exchange.

The concept of commodification can be translated into neoclassical economic terms by interpreting it as sort of a cultural externality (*e.g.,* when sexual prostitution is said to diminish the quality of other people's relationships) or as a lexicographical preference ordering on the social welfare function (so that equality in the distribution of certain goods trumps other allocative or distributional considerations, as in Tobin, 1970). But this interpretation does not do justice to the sense in which those who advocate the concept are concerned with the social construction of perceptions and preferences. Accordingly, the concept is best explained (see, *e.g.,* Kelman, 1981) as relating to potential changes in the content of individual utility functions or of the cumulative social welfare function.

2.5 Legal doctrines regulating freedom of contract

To a considerable extent, the actual legal doctrines restricting freedom of contract can be understood in economic terms, in that most doctrinal restrictions roughly correspond to situations of market failure or high transaction costs, of the sort discussed in the previous subsections. This correspondence is not exact, however; and a survey of the main restrictions will illustrate the roughness of the relationship.

2.5.1 Formal requirements for contracting

The law imposes a number of formal requirements that must be satisfied in order for contractual agreements to be enforceable in court. In general, as discussed in §2.1.1 above, failure to satisfy these requirements results merely in denial of public enforcement rather than any negative sanction. Under US law, for instance, a contract for the sale of goods with market value above $500 (a threshold that has not been adjusted for fifty years) is not ordinarily enforceable in the absence of a writing.[41] Nonetheless, there are no penalties for entering into an oral contract of this sort and in fact such bargains are regularly concluded and performed without incident.

In many other situations, though, denial of public enforcement operates as a binding sanction, especially in one-shot deals where the potential gain or loss from breach of contract is high compared to the value of future business relations. For instance, few businesses would use oral contracts to govern the production and sale of custom-made machinery; and fewer lawyers would advise their use. In such cases, formal requirements can have important regulatory consequences; and their costs can deter exchanges from taking place (as when a business decides not to go public because of the burden of complying with the SEC's disclosure and auditing requirements).

Most legal formalities, like many default terms of the sort discussed above in §2.1.3 are justified, at least in part, by the need for coordination or by administrative needs of the legal system. The rules of offer and acceptance, for instance, discussed below in §3, are typically interpreted as social conventions that serve to help contracting parties ensure that they attach similar meaning to their words and actions and that this meaning will be understood by third parties interested in the agreement. Similarly, the law encourages parties to structure their arrangements and to create and present evidentiary records in a way that lowers the *ex post* costs of contract enforcement—both because of the economies of scale involved in setting up any particular administrative regime, and because it is judged infeasible or politically undesirable to tailor court fees to the individual costs of dispute resolution.

Many formalities, however, also serve regulatory purposes. For example, the requirement that litigation documents be submitted on standard size paper (a formality imposed by virtually all court systems) serves purely administrative purposes, but the requirement that a minor's application for a marriage license be signed by a parent or guardian is also intended to help prevent the minor from making a hasty and irrevocable decision. And some formalities, such as the doctrinal rule that silence in the face of an offer does not ordinarily constitute acceptance, serve both coordinating and regulatory functions.[42] Such a rule indeed comports with standard interpretive conventions, but it also serves to discourage parties from sending purely opportunistic offers in the hopes that the recipient will somehow overlook them.

When viewed as restrictions on contractual freedom, formalities are plausibly justifiable on at least three of the theories of market failure discussed in

[41] UCC 2–201.

[42] See Restatement (Second) of Contracts 69 (1979). In exceptional circumstances, however, such an inference is justified, as when there has been a course of dealing that leads the offeror to expect a response.

§2.3 above. First, formalities can operate to correct informational problems. Parties entering into contracts may not always realize that they are doing so, especially if they are amateurs or newcomers to the relevant commercial community; similarly, they may not understand all of the specific obligations entailed by contracting. Formalities that warn such parties against assuming unintended or unwanted obligations can thus prevent inefficient exchanges as well as undesired distributional transfers from the uninformed to the informed. (On the other hand, as many anti-formalist critics have argued, hidden formalities can also mislead naïve agents into thinking they have entered into binding legal obligations when in reality they have not.)

Second, formalities may provide an effective response to bounded rationality if their presence triggers some cognitive or institutional process that operates as a safeguard against the specific dysfunctional behavior at issue. For example, the federal regulation requiring a "cooling-off" period before completion of a consumer door-to-door sale allows time for additional reflection at a remove from any pressure imposed by the salesperson; and in an organizational setting, writing requirements allow principals more easily to know when an obligation has been undertaken and to monitor their agents for making bad bargains.

Third, formalities can help to correct various negative externalities, especially those that are moderate in magnitude or related to informational asymmetry, by attaching an additional cost to the externality-producing behavior. In markets affected by adverse selection, for instance, formalities that make it harder to disclaim warranties or opt out of medical coverage may keep some high-quality agents from exiting the market, supporting a higher quantity of exchange. Additionally, uncertainty regarding the existence and enforceability of contracts may in some settings adversely affect third parties whose economic fortunes are linked with the contracting agents. Such externalities may explain the traditional common-law rule requiring a writing for contracts for the sale of land or in consideration of marriage—both transactions with potentially significant implications for other members of the contracting parties' family and community. Finally, formalities can help deter rent-seeking by opportunistic agents hoping to take advantage of the informational and behavioral problems discussed just above.

To illustrate these possibilities, we briefly discuss two of the more important formalities imposed by the common law of contracts: the doctrine of consideration, and the statutory requirement that certain contracts be executed in the form of a writing.

The doctrine of consideration. In the US and in other legal regimes that descend from the English common law, contractual promises are not enforceable at law unless supported by what lawyers call consideration. The precise meaning of this concept has long been debated, but its overall import is to require the contract to relate to an exchange that is understood as such by both parties. A promise to make a future gift is not enforceable under the law of contracts, for instance, although it may be possible to achieve a similar result by using some

other legal device such as an equitable trust. While the consideration doctrine has at times been viewed as a substantive limit on contractual freedom, modern commentators (following Holmes, 1897, and Fuller, 1941) view it largely as matter of form, in that it can often (though not always) be avoided by the use of some other formal device such as a special writing, the transfer of a nominal sum, or delivery of a symbolic token. As such, it raises the relative cost of entering into those contracts to which it applies.

Although its grounding in exchange would seem to suggest a close connection between the consideration doctrine and the promotion of economic welfare, the doctrine would seem to diverge from simple efficiency in at least two respects. First, many non-exchange promises also enhance economic welfare. A donor's commitment to make a gift, for example, enables the beneficiary to engage in specific anticipatory investment, thus lowering the donor's cost of providing the beneficiary with any given level of utility (R. Posner, 1977, and Shavell, 1991). Second, the lawyer's understanding of what counts as an exchange is narrower in practice than an economist's would be. In traditional common law, for instance, a promise to guarantee the debt of another person was not enforceable unless the party seeking to enforce could show that the promisor received a specific promise or benefit in exchange for the guaranty; it did not suffice to observe that commercial parties do not typically make such guaranties unless they stand to gain from the extension of credit and are hoping to induce it. Similarly, certain open-ended promises, such as promises to buy or sell goods in a quantity to be specified by the promisee, were traditionally deemed unsupported by consideration, on the theory that the promisee retained discretion to fix terms that would eliminate all benefits received by the promisor.

The doctrine of consideration has long been controversial; and civil law systems, such as those in continental Europe, do not employ it. In its application to donative promises, the doctrine is probably best justified as a response either to bounded rationality or to *ex ante* rent seeking by potential beneficiaries. It is at least plausible that impulsive promises are a greater problem when motivated by generosity rather than self-interest, and requiring more elaborate formalities before such promises become enforceable may be justified in order to protect the interests of the donor or of other potential beneficiaries of his largesse who might fare better upon more thorough deliberation. Similarly, there is an obvious incentive for overreaching by potential beneficiaries, and requiring an additional formal protection such as an equitable trust, which imposes fiduciary duties on the trustee and in practice requires the participation of a legal professional to assure its effectiveness, probably serves to police such opportunism better than a mere writing requirement would.

Writing requirements. While written documents are typically regarded as better guides to the intention of the parties than oral testimony or circumstantial evidence, the general common-law rule is that most contractual promises are enforceable without a writing, if they can be proven to a court's satisfaction. Certain categories of contracts, however, must be evidenced in writing to be

legally enforceable. In common-law systems, such contracts include contracts for the conveyance of interests in real property, contracts in consideration of a party's marriage, third-party guaranty contracts, contracts in which an executor agrees to pay the debts of an estate, contracts that cannot be completed within a year, and, as indicated above, contracts for the sale of goods above a monetary threshold (in the US, $500). These various writing requirements are collectively referred to as the "Statute of Frauds" because they derive originally from a parliamentary statute of 1677 that was justified by its proponents as a safeguard against false claims of contractual agreement. Similar writing requirements have been extended by statute to a variety of other transactions, including contracts that designate property as collateral for a secured loan and contracts for the sale of any personal property with value over $5000; and in some settings, such as option and guaranty contracts, a written promise suffices as a substitute for consideration.

These various requirements surely serve the purposes of *ex post* efficiency in the enforcement setting; for example, courts might well wish to screen out contractual disputes where the key evidence is not just oral but stale (the usual rationale for the one-year provision of the Statute of Frauds). Given that the interpretation of oral evidence is typically more uncertainty, allowing high-value contracts to be alleged on purely oral evidence is particularly costly in terms of incentives for *ex post* rent seeking. And given the considerable private incentives for contracting parties to memorialize their agreement in writing even apart from the prospects of a dispute, the absence of a writing justifies, on Bayesian grounds, the inference that further proceedings are unlikely to be of much value.

Most of these requirements, however, also correspond to some standard type of market failure such as externality or bounded rationality. Guaranty contracts, secured lending, and contracts for the conveyance of land, for instance, all implicate the interests of third parties, and in former times, so did contracts in consideration of marriage. Requiring such transactions to be evidenced in writing reduces information costs for third parties trying to determine the status of assets in which they may also have a claim. Similarly, the expected costs of bounded rationality are especially great in high-value contracts, contracts likely to be entered into under circumstances of emotional stress, and contracts that reach far into the future. In such cases, the extra transaction costs of a formal writing may be justified because they deter impulsive or myopic decisions.

2.5.2 Substantive limitations on contracting

In addition to the practical restrictions imposed by formal requirements for contracting, some agreements are simply unenforceable as a matter of substance. Such limits on enforceability could serve to strike the entire contract, providing an affirmative defense to any liability for breach; or they could strike certain terms of the contract only, leaving the rest of the contract fully enforceable but with the offending terms replaced by default terms, or by terms the court deems fairer (see Craswell, 1993, for a discussion). The major defenses we consider here include fraud, lack of capacity, mistake, duress, undue influence,

unconscionability, and offense to public policy.

Fraud and unilateral mistake. All legal systems refuse to enforce contracts that are based on sufficiently incorrect or asymmetric information, at least in cases where the uninformed party is unaware of his informational disadvantage. Enforcement is especially disfavored when one party to the exchange has caused the other to become misinformed, for example by misrepresenting material facts relating to a proposed exchange. Under the common law, this disfavor is reflected in the defense of fraud.

The fraud defense is justified on efficiency grounds for two reasons. First, it deters inefficient exchanges that would not have taken place but for the fraud; and second, in cases where fraud takes the form of undertaking efforts to deceive others, it discourages rent-seeking. Determining what statements count as fraudulent, however, is not always easy. Vague or ambiguous statement may raise interpretation issues of the sort discussed *infra* in §4; and even statements that are literally true may be interpreted as making an implied fraudulent representation. For economic discussions of this issue, see Ayres and Klass (2005) and Craswell (in press).

Additionally, the defense is not limited to cases of affirmative deception: it can also be asserted when one party withholds critical information that the other reasonably expects to be disclosed. Such fraud by omission is controversial from a legal perspective because of the difficulty of determining when there is a duty to disclose, and from an economic perspective because the duty to disclose information will affect incentives to acquire it.

Similarly, the doctrine of unilateral mistake allows a party to escape a bargain if his assent was based on a significant misapprehension that the other party could have easily corrected (as when a contractor underbids a job because of a calculation error that is evident from the large discrepancy between the mistaken bid and all others received). Here the law takes the position that the small effort required to correct the mistake is justified by the dislocation imposed on the mistaken party and the significant chance that the bargain is inefficient.

In order for doctrines like fraud or unilateral mistake to be welfare-improving in practice, however, courts must have a reliable way to distinguish between asymmetric information on the one hand, and conscious risk-taking in the setting of incomplete but symmetric information on the other, else too many *ex ante* efficient transactions will be avoided. In this area, accordingly, there is potential for divergence between what the law provides and what would be most efficient from a second-best perspective. In general, the courts have been sensitive to such concerns and have limited the scope of the doctrines to fairly significant imbalances of information (and in the case of fraud, have imposed special requirements of proof and pleading).

Incapacity. When one of the parties to a contract is incapable of weighing costs and benefits, there is no longer any basis to presume that the contract

is efficient. Similarly, the law declines to enforce contracts entered into by children, intoxicated persons, and the mentally disabled; this result also deters rent seeking by those who would take advantage of the incompetent. In contrast to the situation of fraud and mistake, however, there are circumstances in which an incompetent person will be held to his bargain. One such circumstance is when the contract is for the sale of "necessities" (*e.g.,* clothing or shelter that are judged to be in the objective interest of the buyer). Another is when the competent party could not have known that the counterparty lacked capacity, although many courts will still supervise the bargain in such cases to ensure its substantive soundness.

The different treatment of the capacity and fraud defenses is justified by the differing costs of overcoming the specific transactional failure at issue. In the case of fraud, the failure can be overcome at no cost by abstaining from misrepresentation, and in the case of unilateral mistake, the failure can be overcome at low cost by correcting the mistake. In cases of incapacity, however, it may not be possible to correct the problem without losing the transaction entirely, so the law allows plainly efficient exchanges (or in the case whether the incapacity is hidden, exchanges where it is plain that no rent-seeking took place.)

Mutual mistake. In some cases, the law refuses to enforce bargains in which the parties' *ex ante* beliefs were sufficiently different from the state of the world, even when those beliefs are the same and when neither party could easily have corrected the error. For example, in the celebrated case of *Sherwood v. Walker*, the court allowed the seller of a pregnant cow to void the deal on the grounds that, at the time of the bargain, both parties mistakenly thought that the cow was barren and set a price that corresponded to its slaughterhouse value. This doctrine, known as the defense of mutual mistake, is subject to much confusion and there is no canonical economic explanation of it. The legal confusion results because it is difficult in practice to know whether the parties were truly mistaken or just engaged in a hedging transaction. From the economic viewpoint, Posner and Rosenfield (1977) view the doctrine as substituting, in a situation of costly contracting, for a complete contingent contract in which it is efficient to call off the sale in the state of the world where the mistake is discovered. Ayres and Rasmusen (1993), on the other hand, argue that the mutual mistake doctrine, in contrast to unilateral mistake, undermines incentives for information acquisition, and conclude that it is dominated by a suitably limited unilateral mistake doctrine. Indeed, some legal commentators (*e.g.,* Chirelstein, 2006) suggest that in practice the mutual mistake doctrine is used when the court suspects fraud or unilateral mistake but cannot clearly make out the elements of those doctrines under the facts. The doctrine, accordingly, may afford a good opportunity for further theoretical investigation.

Duress and restrictions on contractual modifications. The doctrine of duress denies enforcement to contracts formed under conditions of unacceptable coercion; the standard hypothetical is a contract signed at gunpoint. The

key to the doctrine, of course, consists of which conditions are deemed unacceptable. Some courts and commentators have held that difficult economic circumstances can amount to duress; while others suggest that hard bargaining is permitted unless the beneficiary of the promise is responsible for the difficult situation (for example by having isolated the promisor from other potential contract partners) or owes some duty with regard to the difficult situation (for example, by having been granted a local monopoly). Whether a contract is void for duress, however, depends not just on external conditions but also on the content of the contract. For instance, a contract to salvage the cargo of a sinking ship at 10% of the value of the cargo might be enforced, while one that set a price of 90% of cargo value probably would not.

A similar functional problem arises in contracts that modify duties created by previous contracts, because many modifications are entered into under circumstances where one of the parties is dissatisfied with the original contract and is threatening to breach it. For example, in the case of *Alaska Packers v. Domenico*, the court held unenforceable a modification that provided a 66% increase in wages to a crew of sailors who threatened to strike on short notice in remote waters.

Much of the legal commentary on this doctrine justifies it on the norm of party autonomy, on the theory that a choice made under coercive circumstances is no choice at all. This justification is difficult to square with the economic viewpoint, however, unless one interprets it as claiming that the harsh circumstances somehow interfere with the exercise of rationality. But in most cases in which the doctrine is applied, rationality does not seem at issue—even in the gunpoint hypothetical, the victim acts rationally in choosing life over money. For this reason, some have argued (*e.g.,* Bar-Gill and Ben-Shahar, 2004) that the doctrine be made unavailable in circumstances where the coercive threat is credible—that is, where the party making it would actually find it in its interest to carry it out if the coerced party does not agree to the exchange in question.

A better economic justification of the doctrine is found in the phenomenon of rent seeking (*e.g.,* Tullock, 2005)—that we wish to discourage investments in coercion or against coercion. In the gunman hypothetical, the rent seeking is obvious, but as Cooter (1982) and Cooter et al. (1982) suggest, a similar argument can apply to ordinary hard bargaining in situations of bilateral monopoly. Even when an exchange is efficient, in the absence of a well-defined mechanism for dividing the gains from trade, the parties may destroy part of the surplus in attempting to influence its distribution.[43] Additionally, in dynamic settings, the prospect of earning such rents may lead to overinvestment or excess entry (cf. Mankiw and Whinston, 1986). Thus, by reducing the set of possible threats, the law can, in principle, limit the available rents and encourage a more efficient equilibrium.

The rent-seeking explanation is especially apposite in the case of contrac-

[43] Although when the bargaining is conducted under symmetric information it is difficult from a game-theoretic perspective to see why surplus would be destroyed (see Hermalin and Katz, 1993, and the discussion in §2.2.2).

tual modifications, because parties who make specific investments in an initial contract become vulnerable to holdup. In many circumstances, parties will be reluctant to make specific investments in settings in which contract renegotiation contract is possible; accordingly, it is, in principle, beneficial for them to commit not to renegotiate if they can credibly do so (Aghion et al., 1994).[44] The law of contracts has been unwilling to allow parties to rule out modification, however, perhaps because of concerns for imperfect information and bounded rationality (Jolls, 1997). It has, however, used various strategies over the years to limit modifications that appear motivated by opportunism. In particularly egregious cases, as Shavell (2006) points out, courts will void the attempted modification on grounds of duress or unconscionability. In less extreme situations, more flexible tools are available. Traditionally, such regulation operated under the formal rubric of the consideration doctrine, on the theory that a modification that unilaterally altered the contract in one party's favor was not a true exchange. The modern approach, however (*e.g.,* Hillman, 1979) requires the modification to pass a substantive test of fairness or good faith. For instance, a modification made in proportionate response to new circumstances, unanticipated at the time of contracting, would generally be enforced, while an outright attempt to rewrite the original terms of the bargain would not. Because it is difficult to identify objective criteria for which circumstances are unanticipated and what sort of adjustment would be proportionate, however, the standard of enforcement is uncertain, so that some incentives for rent seeking do remain.

Undue influence. A fourth standard limitation on contractual freedom is undue influence. This defense applies when one party is deemed to have such influence over the other that the latter lacks the requisite autonomy for contracting. Classic examples include contracts between lawyer and client, physician and patient, caregiver and invalid, and, traditionally, husband and wife.

The undue influence doctrine overlaps in function with the other substantive doctrines listed above. It shares with incapacity the element of bounded rationality, in that the influential relationship may be thought to interfere with the vulnerable party's ability to exercise independent judgment (and as with incapacity, undue influence does not entirely bar the enforcement of a contract, but rather subjects its terms to *ex post* regulation on grounds of substantive rationality and distributional fairness). It also shares with fraud the element of asymmetric information (in that the relationship entails a reposition of trust), and with duress the element of rent seeking (in that assent may be given in response to an implicit threat to withdraw the relationship and its associated benefits).

Undue influence is less commonly invoked in the context of person-to-person contracts than the other doctrines discussed above, but it operates as a cornerstone of the law of fiduciary relationships, and thus of much law governing professional relationships and contracts between organizations and their governing agents. For example, the corporate law duty of loyalty, which subjects

[44] Shavell (2005) notes that there is no empirical evidence that parties are able to do so.

dealings between corporations and their officers or directors to special scrutiny, is usefully understood as an application of this basic contractual doctrine.

Unconsionability and public policy. Finally, the unconscionability doctrine serves as a sort of all-purpose limitation on contractual freedom applied *ex post* in cases where the court deems the parties' exchange to be sufficiently unfair in either substantive or procedural terms. Some commentators (*e.g.,* Epstein, 1975) have suggested that the doctrine can be justified in economic terms to the extent that it operates as a discretionary extension of the other major affirmative defenses in marginal cases. For instance, the doctrine has been used to protect consumers or unsophisticated small business people from contractual terms that imposed large *ex post* risks and that were buried in fine print or otherwise insufficiently disclosed, which is in some ways analogous to fraud or incapacity. (As noted in the earlier discussion of fraud and non-disclosure, it is not always easy to determine what kinds of statements or omissions should be treated as fraudulent.) The doctrine may also be partially justifiable on the economic grounds of externality (*e.g.,* E. Posner, 1995), to the extent that it prevents contractual creditors from driving their debtors into insolvency, and thus imposing financial obligations on the debtors' families or on the public. Similarly, when viewed in connection with a variety of legal rules that protect debtors in extreme situations, such as bankruptcy law and other statutory limitations on debt collection, the doctrine can be understood as responding to failures in the markets for credit and insurance.

These efficiency rationales cannot entirely account for the actual operation of the doctrine, which sometimes appears merely to reflect the courts' reluctance to hold a hapless party to a bad bargain, or a paternalist concern that a party has undertaken an excessively high risk. The doctrine's implications for freedom of contract, however, should not be overstated, despite its prominent treatment in academic discussions and introductory student texts. Its effect is limited by the requirement that it can only be applied by a judge as opposed to a jury, and by most judges' appreciation that, at least in commercial cases, parties have good reason to take risks and adequate opportunity to obtain insurance. As a practical restriction on the parties' ability to get their bargains enforced in court, accordingly, it is probably rather less important than either the costs of litigation or than the more general psychological tendency of both judges and juries to resolve ambiguous facts in favor of the party with whom they feel the strongest sympathy.

Similarly, courts sometimes decline to enforce contracts on the grounds of public policy, although this concept is less a specific doctrine than a label for a general practice of limiting contractual freedom on a variety of rationales. The main difference between unconscionability and public policy is that the former is based on unfairness, while the latter is based on the idea that enforcement runs counter to the general goals of the legal framework or judicial system, even though no specific statutory or common law rule contains an explicit ban on the contract in question. For example, contracts to engage in gambling or

in ostensibly immoral behavior were traditionally denied enforcement on such grounds, as were contracts in restraint of trade and contracts thought to promote litigation. A prominent modern case applying the concept is the Baby M case, in which the New Jersey Supreme Court refused to recognize a contract which purported to transfer parental rights in exchange for financial consideration. Because of the breadth of this category, probably the best way to summarize it is simply to say that while the principle of freedom of contract is generally respected by the legal system, it remains subject to a vague and implicit set of limits that operate at its margins, on analogy to the limits imposed by other fields such as tort and criminal law. A more extensive discussion of the concept can be found in Farnsworth (2004).

3 Formation of Contracts

In many standardized or spot market transactions, parties enter into contracts without any significant amount of bargaining. In more complicated situations, however, a contract typically follows an extended set of communications that can include offers, counter-offers, and other exchanges of information. Such communications are governed by an elaborate framework of legal rules that determine when, how, and on what terms contractual obligations are created.

Compared to other areas of contract law such as the rules governing remedies for breach, there has been relatively little discussion of this doctrinal subject from an economic point of view. In part, this neglect results from the fact that these doctrines are esoteric and little known outside the legal profession. Additionally, legal discussions on the topic tend to focus on the narrow problem of channeling the parties' communications into conventionally recognizable forms, rather than on more direct regulatory concerns.

From an economic viewpoint, however, the law of contract formation is relevant because it influences the outcome of exchange. By attaching consequences to the various acts and omissions that individual bargainers can choose from in a negotiation, legal rules affect the parties' incentives to make and to respond to offers, to exchange information, and to communicate with one another at all.

In this section, accordingly, we survey the various dimensions in which pre-contractual behavior can affect the efficiency of exchange, and the way in which the law of contract formation affects incentives along those dimensions.

3.1 Pre-contractual behavior

Even before they meet, potential contracting parties make a number of economic decisions that influence the possible gains from future trade. Not all of these decisions, however, will necessarily be made optimally. Some decisions create positive or negative externalities, and others entail relational investments that are vulnerable to holdup. In this subsection, we identify and discuss three overlapping dimensions of pre-contractual behavior: searching for contractual partners, investigating the value of exchange, and making pre-contractual investments.

3.1.1 Searching for contractual partners

In order for parties to be in position to enter into exchange, they must undertake the expense of searching for potential trading partners. Such expenses include advertising, correspondence, travel, and the parties' time. From the perspective of a social planner, one would want the parties to undertake such efforts up to the point where the marginal costs of additional search just outweigh its expected marginal value. While search can be modeled in various ways (see, *e.g.,* Diamond, 1987), under plausible assumptions, the value of an additional unit of search lessens as the quality of the bargain in hand increases. It follows that it is, in general, optimal for a party to search up to some cut-off level of satisfaction, and then stop. This cut-off level will depend on the cost of search, the distribution of information, the expected bargaining game to be played once search is completed, and so on.

The level of search that is privately profitable for an individual buyer or seller, however, is not necessarily the same as the level that would be socially optimal, for two reasons. First, in markets with bilateral search, each person's search efforts provide a positive externality that reduces the search costs of others. Second, to the extent that there are economic rents associated with trading (*i.e.,* if parties buy or sell at prices that diverge from their reservation prices), some amount of search is motivated by the desire to find a better distributional outcome. These effects work in opposite directions, so it is difficult to generalize about what public policies would be optimal in this regard, but it is possible in principle that enlightened regulation could improve social welfare.

For example, §3.3 below explains how various legal doctrines operate to promote the early formation of contractual liability. Such doctrines will thereby affect the level of search, but the direction of the effect is ambiguous. From the *ex post* perspective of parties who have already found each other, such rules should reduce the incentive for additional search, because a party who holds a binding commitment has less need to search; and a party who is bound faces reduced value from search because even if she finds a better bargain, she will still be liable for the first one. But from an *ex ante* perspective, parties may be more willing to undertake the cost of search if they are assured that any trading partners they find will stick with their deal. Which effect is more important depends on the details of the situation.

3.1.2 Acquiring and disclosing information about the value of exchange

In order to conduct exchange, the parties not only must find each other, but they must also determine whether trade is worthwhile. Both the acquisition and disclosure of such information can be costly, and the parties' willingness to incur these costs will depend on whether they can be recouped in subsequent

contract bargaining and performance.

As with informational investment more generally, imposing a duty to share information tends to undercut the incentive to acquire it, so there is a potential tradeoff between efficient production of information *ex ante* and efficient use of information *ex post*. In this regard, Kronman (1978a) argued that requiring commodity traders to disclose private information about market value would undercut their incentives to do market analyses and make forecasts; and Matthews and Postlethwaite (1985) showed how requiring manufacturers to disclose product quality could discourage product testing.

Whether information should be produced, of course, depends upon whether it is socially valuable, as opposed to simply having private redistributive value. For that reason, if information acquisition is, on balance, socially wasteful, it is best to impose a disclosure requirement in order to deter rent-seeking (or a tax on acquisition efforts or windfall profits if that is administratively cheaper).

Whether information is acquired or disclosed before exchange, however, will depend not just on regulatory requirements, but also on property rights, the relational nature of the information, and market structure. For example, in Shavell's (1994) model of information acquisition, a party with exclusive rights over a particular item of property has efficient incentives with regard to acquiring information that bears on the common value of that property, because he internalizes that common value when he sells or uses the property. One could imagine circumstances, however, in which a party without formal property rights had similarly good incentives because he held a monopoly position in bargaining and because the information increased his private value for the property, but not the owner's.

Disclosure may also be compelled by the structure of expectations in bargaining. For example, Grossman (1981) presents an influential model of product warranties in which buyers rationally assume the worst about any seller who does not disclose its private information. Sellers with relatively good information are thus led to disclose it in order to avoid the effect of buyer skepticism, and the result is a separating equilibrium in which the private information is revealed.[45] Note that disclosure induced by such game-theoretic incentives can discourage information acquisition every bit as much as disclosure required by legal regulation.

Even in cases where mandatory disclosure would not adversely affect the incentive to acquire information, it is still necessary to compare the benefits of disclosure against the costs. In many contexts, especially those involving ordinary consumers, the costs of disclosure will include communication costs—for example, the time that it takes buyers to read and process the information; or

[45]To be precise, information unravels to the point where the cost of credible disclosure equals the difference between the price charged by the pooled low-quality sellers and the price that the highest-quality member of this pool could obtain through disclosure. Observe the unraveling result depends on the disclosure being effective to convey information. See, *e.g.,* Fishman and Hagerty (2003), who show that if the fraction of customers who can understand a disclosure is too low, then mandatory disclosure benefits informed customers and harms the seller.

the potential cost of being distracted from other, more important information. These communication costs have rarely been incorporated into formal economic models, though Bebchuk and Shavell (1991) is an exception. Craswell (in press) provides an informal discussion of these costs.

Finally, it must be noted that in some instances *partial* disclosure of information is worse than no disclosure (Hermalin and Katz, in press). This is true when partial disclosure facilitates inefficient discrimination. For instance, disclosure of information correlated with on-the-job performance (*e.g.,* health status) could depress the wages of those disclosing the correlate in equilibrium, causing some talented individuals to exit the labor market.

3.1.3 Pre-contractual investment

The previous discussion emphasized that investigation of exchange value can operate as a sunk investment that is vulnerable to rent-seeking in later bargaining. The same is true of pre-contractual investments generally.

Much of the law of contracts is designed to protect investment incentives, and §4 and §5 below discuss the effect of interpretation rules and contract remedies on those incentives. But these rules come into play only after a contract has been formed. Some investments, however, must be undertaken before it is practical for contractual liability to attach. For example, search efforts inherently must precede contractual negotiation (else who would one negotiate with?), and negotiation must precede the formation of a contract. But both search and negotiation are costly, and their costs are at least in part relation-specific. Similarly, other investments can be delayed until negotiations are completed, but their value is reduced in doing so (Katz, 1996c). Suppliers of goods, for instance, can typically reduce their production costs by buying materials when prices are low or by doing advance work when business is slow; and buyers can increase their value from exchange by investing in complementary inputs. But if they wait until they are finished bargaining to begin preparations, many such opportunities will be lost.

It is not desirable to provide complete protection for pre-contractual investments, conversely, because such protection would lead to excessive reliance. As the parties negotiate, they may discover information that reveals that the intended exchange should not be pursued. If this happens, any relation-specific investments will be wasted. Optimally, the parties should take the risk of wasted investment into account before making them. The rules governing contract formation, accordingly, should ideally be designed to promote optimal reliance at the optimal time, balancing the benefits of productive investment against the costs of waste.[46]

[46]The mechanism for paying compensation for a taking (governmental exercise of its right of eminent domain) must strike a similar balance (Hermalin, 1995).

3.1.4 Strategic behavior in bargaining

In §3.1.2 above, we discussed incentives to produce and share information regarding the value of exchange. But information exchange is not the only dimension of efficiency in bargaining. Negotiating parties can dissipate resources through various types of rent-seeking: by excessive communications, by haggling and stalling, and by hard bidding that risks losing the bargain entirely. One way to deter such behavior is through substantive legal doctrines that limit the kinds of bargains that can be enforced, and thus lessen the temptation for overreaching; for example, Cooter (1982) justifies the doctrine of duress on such grounds. But another way is to create contract formation doctrines that impose the cost of lost bargains on parties who cause them through excessive rent-seeking.

In order to address the problem of excessively hard bargaining, however, it is necessary to develop a clearer understanding of incentives to engage in it. As we discussed in §2.2.2 above, in the *absence* of asymmetric information the parties should be able to reach an efficient bargaining outcome without outside intervention.

Symmetric information is likely the exception rather than rule, however. Complete symmetry assumes symmetry of knowledge about preferences, among other parameters, which is typically not true (see *e.g.,* Farrell, 1987b). If parties bargain under asymmetric information it is well-established (*e.g.,* Samuelson, 1985; Farrell, 1987b; Schweizer, 1988) that they will often fail to reach a first-best outcome. The theory of mechanism design implies that for any asymmetric information bargaining game, there is, in principle, some mechanism that maximizes the parties' expected bargaining surplus subject to their participation and truth-telling constraints; that is, achieves the second best.[47] But any particular bargaining game may not be the optimal mechanism, so the ultimate efficiency of the bargain could depend on limitations on bargaining imposed by the law on contract formation.

3.2 Avoiding miscommunication

Finally, before we move to a discussion of legal doctrine, we consider the important role played by formation rules in avoiding miscommunication. Negotiating parties need to know when bargaining has been completed and whether a binding obligation has been formed, since if they have inconsistent understandings in this regard, significant value can be wasted. If a party believes that he has entered into a binding obligation when in fact he has not, he may waste resources attempting to perform or may turn down other profitable opportunities. Conversely, if a party is unaware that a contract has been formed, she may fail to make adequate preparations or even incur multiple liability, resulting in breach of contract and associated losses. One of the most important functions of contract formation law, accordingly—and the function to which legal scholars have probably devoted the most attention—is to promote clear channels of communication so that the parties may know where they stand.

[47] See Laffont and Martimont (2002), Bolton and Dewatripont (2005), or Caillaud and Hermalin (2000b) for introductions to mechanism design.

The law of contract formation pursues this task in two distinct ways. First, it establishes authoritative forms and terms of art that the parties can use when negotiating their agreements. Second, it allocates the risk of miscommunication so as to encourage parties to take optimal precautions to prevent and insure against misunderstandings, in the same way that the law of torts allocates liability for accidents to the least-cost avoider and least-cost insurer. For example, the doctrines of unilateral and mutual mistake, discussed in §2.5.2 above, illustrate how loss allocation rules can strengthen the parties' incentives to avoid misunderstandings and to allocate any remaining risk on their own.

This issue of clarity in communication is most usefully understood in terms of the problem of interpretation, which we discuss at length in §4 below. Determining whether a contract has been formed raises most of the same issues as determining the terms on which it has been formed. For example, one must identify the meaning of phrases that are used in negotiation, reconcile interpretative differences among the parties or between the parties and the court, establish default rules that apply when the parties have left their negotiations open or spoken ambiguously, and so on. We defer consideration of such issues until the next section. Keep in mind, however, the connection between the legal doctrines discussed here and their role in interpretation discussed in §4. For instance, as we discuss in §4.4.2 below, pre-contractual communications often bear on the interpretation of a contract, especially with regard to issues where its text is silent or unclear.

3.3 Legal doctrines addressing contract formation

The body of legal doctrine relating to contract formation is particularly elaborate, and we do not attempt to survey it here. Instead we discuss a number of key doctrines that help address the central economic aspects of pre-contractual behavior: disclosure, investment, and efficiency in negotiation.

3.3.1 Offer and acceptance

The classic distinction between contract law and other sources of legal obligation is that contract is grounded in voluntary agreement. The fundamental principle of contract formation, it follows, is that there must be mutual assent in order to establish a binding contractual obligation. Assent refers not, however, to the parties' private mental states, which clearly cannot form the basis of a public system of enforcement, but to our inferences regarding those states as drawn from the parties' actions and statements.

To simplify this problem of inference, lawyers often say that to form an agreement there must be an offer (which evidences one party's assent) and also an acceptance (which evidences the counterparty's assent); and thus this body of doctrine is often referred to as the law of offer and acceptance. It should be recognized that this way of describing the matter is a formalism that allows

interpreters to focus on certain stereotypical features of negotiating a contract and abstract from others. In actual contexts it is often difficult or artificial to identify a particular communication that counts as an offer or an acceptance, and some doctrinal systems (such as the law of sales under UCC Article 2) dispense with the effort. In general, however, the traditional common law has followed a highly formalistic approach in this regard, with the result that the treatises and casebooks are filled with a variety of arcane and colorfully named mechanical rules. Such rules include, among others, the "mailbox rule" (*i.e.,* when parties communicate a distance, a contract is formed at the moment an acceptance is dispatched, not when it is received), the "mirror-image rule" (in order to constitute an acceptance, a responding communication must mirror the offer in all relevant respects, else it is deemed a rejection and counter-offer), and the "last-shot rule" (if parties exchange a series of non-matching offers followed by one of the parties commencing performance, a contract is deemed to be formed with the beginning of the performance constituting the acceptance, and with the terms of the contract supplied by the final offer outstanding prior to acceptance). Some of the formalistic rules have received substantial economic attention (see, *e.g.,* Baird and Weisberg, 1982, and Ben-Shahar, in press, on the last-shot and mirror-image rules), but many others have not.

Over the last half century, such formalistic rules have come under increasing criticism, and more recent developments tend to de-emphasize formality, making it easier to establish liability without complying with prevailing formal conventions. This movement away from formality has had its most important effect on communicative efficiency; and is discussed at greater length in §4.4.1 below. Similarly, and simultaneously, the trend has been for the courts to establish that liability is incurred earlier in the bargaining process (with the mailbox rule providing an early precursor in this regard). This trend is probably most important with regard to incentives for pre-contractual investment, and we take it up in the subsection immediately following this one.

In addition, however, the particular rules of offer and acceptance also influence the transactions costs of negotiation in potentially significant ways. As an illustration, consider the doctrine of silent acceptance (Katz, 1990a, 1993), under which an offeree must respond affirmatively in order to create a binding obligation; silence operates as an acceptance only in special circumstances. The rule can serve to conserve on message costs. Under it, two messages are required to form a contract, while only one is needed in negotiations that do not result in exchange. A converse rule under which an offeree must respond in order to avoid being bound, would require one communication for a contract and two for a rejection. In the usual context where the majority of offers are rejected and rejections and acceptances cost the same to transmit, the common-law rule is efficient. But in settings where acceptance is likely or where acceptances are costlier to transmit than rejections (for instance, because of a mirror image requirement), the rule is inefficient.

Additionally, the common law rule reduces the costs of rent-seeking through opportunistic offers. Under a silent acceptance regime, offerors will have an incentive to propose inefficient exchanges to offerees with high response costs, in

the hopes that the offeree will choose to accept a mildly unprofitable contract in preference to incurring the costs of sending rejections. Such incentives explain why negative option plans of the sort offered by mail-order book clubs and telephone companies typically provide buyers with substantial up-front benefits as an inducement to enter the plan.

Finally, based on an extensive survey of the case law, Craswell (1996b) argues that offer and acceptance doctrines are widely used by courts to promote efficient relational investment. In his view, courts are often in position to evaluate *ex post* whether reliance has been efficient, and in such cases, they can provide good incentives by finding a contract when a party relied reasonably in response to a pre-contractual communication, but finding no contract when the party relied either excessively or not at all. Whether Craswell's argument provides a normative justification for courts practices in this regard, however, as opposed to an accurate positive account of what they do in fact, is open to dispute. In the first place, his survey of cases suggest that the fact of reliance is decisive to a finding of liability only in relatively close cases, not in all cases, so the incentive effects of this practice may be limited. Additionally, in order for courts to employ this tool, they must be able to judge the efficiency of reliance in hindsight. The practical limitations of the legal process, the fact that reliance decisions frequently entail nonverifiable actions, and the cognitive biases associated with hindsight (Rachlinski, 1998) may call this assumption into question. Instead, courts may do better to follow the simpler approach of identifying and holding liable the party who is the least-cost avoider for wasted reliance. A model based on this possibility is presented in the following subsection.

3.3.2 Promissory estoppel and analogous doctrines

The general principle of estoppel was developed in the English equity courts as a corrective device to preclude the operation of legal rules that were thought to yield an unjust result in specific cases. The more specific doctrine of promissory estoppel gained influence in the 19th century as a way to soften the formalities of the consideration doctrine, but in the latter half of the 20th century in the United States, it became increasingly used to loosen the formal requirements of offer and acceptance.

The key element necessary to invoke promissory estoppel is relation-specific investment, which lawyers call reliance. Under it, a promisor is precluded from asserting a lack of mutual assent if she made a promise that reasonably can be expected to induce the promisee's reliance, which actually does induce such reliance, and which will result in injustice if not enforced. "Injustice" is admittedly subjective, but in effect it serves to protect reliance that most people would consider reasonable. Excessive or unreasonable reliance is not protected by the doctrine, although there are obvious difficulties in determining reasonableness in this setting. For example, in the case of *Drennan v. Star Paving*, a general contractor used a subcontractor's offer in calculating his bid on a construction job. After the general contractor won the construction contract, the subcontractor attempted to withdraw the original offer on grounds of miscalculation, arguing

that under prevailing rules of offer and acceptance, the offer was revocable until the moment of acceptance. The court held that the subcontractor's bid was non-retractable, not withstanding that there was never any explicit promise to hold it open and that the general contractor could have explicitly bargained for a binding option.

The efficiency of this result, and of the estoppel doctrine generally, depends on whether such liability is necessary to protect investment incentives. Katz (1996c) offers a model of pre-contractual investment in which one party makes an offer and the counterparty chooses to rely; and either of these decisions can be taken over a period of time leading up to the deadline for ultimate performance. Excessively early reliance is inefficient in this model because there is too high a chance that the investment will be wasted; excessively late reliance is inefficient because the productive surplus from investment cannot be enjoyed. The main conclusion of the model is that estoppel liability is desirable if the relying party cannot protect the value of its reliance investment in post-reliance bargaining, but undesirable if it can. The underlying intuition is that if the offering party holds the *ex post* bargaining power, the relying party will refuse to invest unless protected by liability; while if the relying party holds the *ex post* bargaining power, the offering party will refuse to specify its needs for fear of being held liable in circumstances where it does not pay to complete the contract, and being vulnerable to rent-seeking through a last-minute counter-offer in circumstances where it does pay to complete the contract.

Katz's model focuses on the timing of reliance investment, but its underlying intuition also applies to situations in which the extent and type of reliance is in question. As a general matter, reliance investments might be protected through liability rules or by *ex post* bargaining; and in cases where bargaining power is sufficient to provide incentives, legal liability is superfluous and may even provide excessive protection. How these considerations play out in individual bargaining contexts, however, depends on the precise nature and sequence of play. For instance, Bebchuk and Ben-Shahar (2001) consider the possibility of legal rules under which courts can condition liability on the level of reliance investment, on its reasonableness, or on the reasonableness of positions taken in *ex post* bargaining, and show that such rules can be used to promote efficient bilateral incentives. They also discuss the effects of liability rules on the parties' incentives to enter into negotiations initially, and show that prospect of liability need not deter initial negotiation (although under some conditions, it might). The informational requirements of such rules are of course significant, which limits their practical applicability in many situations.

The idea of conditioning liability on the reasonableness of bargaining behavior, however, underlies a related legal doctrine: the duty to bargain in good faith. This doctrine is much less widely used than promissory estoppel; it is applicable when the parties enter into an agreement that is too indefinite for a court to enforce as written, but that could be enforced if the parties were to undertake an additional round of bargaining in which they filled in enough gaps. For example, in *Teachers Insurance & Annuity Co. v. Tribune Co.*, 670 F. Supp. 491 (SDNY 1987), the defendant refused to complete a complicated real

estate deal, following significant negotiations resulting in a commitment letter that referred to a "binding agreement," after learning that interest rates had significantly dropped and that the deal would probably not qualify for favorable tax treatment. The court held that even though the defendant had reserved a further right of approval on the part of its board of directors, it was obliged to continue negotiations in good faith, which in the specific context meant that it could not condition its approval on a contingency (*i.e.,* favorable tax treatment) that was deliberately not included in the commitment letter, and could not withdraw from the deal simply because interest rate changes had rendered it unprofitable.

The precise contours of the duty to bargain in good faith are not entirely clear. For instance, it is difficult to predict whether a court would hold it bad faith for a party to insist that all open terms be resolved in its favor or to ask that some previously settled term be re-opened, on the grounds that the deal had turned sufficiently sour that a fair distribution of rents justified such an adjustment. (Cases applying the modern doctrine of modification, which also turns on a good faith standard, have often held such adjustments to be acceptable.) The good faith doctrine has been criticized for its uncertainty and apparent subjectiveness in application. Nonetheless, the bulk of commentators continue to approve of the doctrine as a means of protecting relational investments in negotiation. In *TIAA v. Tribune*, for instance, the parties had already incurred significant costs and, more importantly, a deal of such complexity could not have been concluded without sinking such costs. Absent some form of pre-contractual liability, it would be difficult to enter into complicated contracts without subjecting reliance investments to the risk of holdup.

Other legal systems have developed analogous rules to deal with this functional problem. For example, the civil law regimes of continental Europe do not make use of the doctrine of estoppel, but they do include an invigorated version of the duty to bargain in good faith, known as *culpa in contrahendo*. Under this doctrine, parties entering into contract negotiations are held to a mutual duty of care that is intended to protect the reasonable expectations with which they enter bargaining. This duty has traditionally been regarded by comparative lawyers as being more demanding than the implicit duties imposed by the doctrine of promissory estoppel, but one can argue that modern development in the common law have effectively brought the two systems closer together in this regard.

3.3.3 Duty to disclose

Finally, the common law also imposes an ill-defined duty to disclose especially important information when negotiating with contractual partners. The duty does not extend to all information a counterparty might find relevant; rather, it is limited to situations in which nondisclosure would be regarded as effectively equivalent to a representation that the information does not exist. For example, in the classic case of *Laidlaw v. Organ*, a trader with advance knowledge of a peace treaty that would shortly result in the lifting of naval blockade bought

up tobacco that skyrocketed in value when the news became public. When the seller sought to avoid the contract on the grounds of fraud, the court suggested that, under the circumstances, the parties were entitled to hold each other at arms' length and to retain the benefits of such information for their private use. On the other hand, a homeowner who sells a termite-infested house to a buyer who is known to be unaware of any problem may well be under a duty to disclose it, especially if the buyer would not be able to discover the infestation though the exercise of ordinary diligence.

The duty to disclose has received substantial attention in the scholarly literature; and it was recognized early on that nondisclosure might have effects on the fairness and efficiency of exchange. In the law and economics literature, it has been recognized, at least since Kronman (1978b), that imposing a duty to disclose could deter information acquisition. Though later scholarly efforts (see, *e.g.,* the discussion in §3.1.2 above) have refined economic understanding of the issue, doctrinal application of these efforts has been limited. Kronman argued that, at minimum, there should be a duty to disclose information that was acquired without effort or expenditure, on the grounds that such a duty would not reduce the availability of information and would improve the *ex post* efficiency of exchange, but even this suggestion has not made its way into the case law, which has tended in this area to focus more on the question of rights than on efficiency.[48] Still, the trend in legal decisions seems to run in the direction of increased disclosure requirements, especially in settings where the interests of consumers, workers, or small businesses are involved.

3.3.4 Overall assessment of the law of contract formation

The division between the law of contract formation and other parts of contract law is not clear-cut. For instance, some of the rules discussed above in §2.5 on freedom of contract could alternatively be interpreted and analyzed as formation rules. Take for example the Statute of Frauds (see §2.5.1). From a freedom-of-contract perspective, the Statute helps to deter boundedly rational parties from entering into contracts without adequate deliberation, and may help internalize costs that are imposed on the public legal system when parties enter into agreements without adequate evidentiary backing. But it also serves some of the purposes discussed above, such as the disclosure of information or the reduction of miscommunication. Similarly, the mistake doctrine allows inefficient contracts to be avoided, but it also encourages parties to share information that would prevent misunderstandings and the wasteful investments that can result from them.

Additionally, as elaborated in §5 below, the rules governing contract formation significantly interact with the rules governing the remedies for breach. For instance, as we will see, promissory estoppel may substitute for a formal offer and acceptance in the appropriate case, but the damages theoretically available

[48] While *complete* disclosure will generally enhance efficiency, *partial* disclosure can reduce welfare (Hermalin and Katz, in press).

to the disappointed promisee may be rather less than the damages that would be awarded if the proper formalities have been observed (although there is scholarly controversy regarding whether courts actually follow this theoretical distinction in practice). Again, these doctrinal complications are likely to provide fruitful opportunity for additional economic research in future years, as is the law of contract formation generally.

4 Interpretation of Contracts: Contractual Incompleteness

Offer and acceptance only begins the process of contractual exchange. In order for the transaction to be completed, the contract must be performed, and if performance is not forthcoming, enforced. Probably the most common source of contractual disputes is differences in interpretation, if only because the parties have limited incentive to pursue a dispute if they can foresee and agree upon its likely outcome. The problem of contract interpretation thus provides a central backdrop for the law of contracts, which contains many rules and principles that are designed to address it.

The legal issue of interpretation corresponds to the economic issue of contractual incompleteness—a topic that has been a central focus of research in microeconomic theory in recent years (*e.g.,* Grossman and Hart, 1986; Hart, 1987; Hart and Moore, 1999; Tirole, 1999). Contractual incompleteness captures the idea that real-life contracting can fail to produce contracts that are as precise and detailed as traditional—albeit possibly naïve—economic theory predicts. Economists typically attribute this failure to an informational asymmetry: the parties to the contract anticipate observing events that they might wish to be contingencies, but which cannot serve as contingencies because they are not verifiable (*i.e.,* observable by a third-party adjudicator of any contract dispute); in other words, the parties will be better informed about payoff-relevant information than any third party. For their part, lawyers often attribute the failure to complete contracts to the inevitable ambiguities in ordinary language or to some bounded rationality on the part of the parties (perhaps arising from their decision not to employ lawyers in the drafting of the contract).

In this section, accordingly, we discuss the problem of contractual incompleteness and relate it to the question of interpretation and to associated legal doctrines. We begin in the next subsection by considering various definitions of contractual incompleteness. Subsection 4.2 then discusses the sources of contractual incompleteness, including *ex ante* determinants present at the outset of contracting, such as bounded rationality and asymmetric information, as well as *ex post* factors such as verification costs and dynamic incentives arising from the prospect of renegotiation. Subsection 4.3 addresses the consequences of contractual incompleteness for the efficiency of exchange; and subsection 4.4 outlines and analyzes the main legal doctrines that govern in this area.

4.1 Modeling incomplete contracts

From a theoretical perspective, it is useful to model a contract as a mapping from *verifiable* events to outcomes. For instance, an insurance contract could contain a provision that related damage to one's car (a verifiable event) to a payment to the insured (an outcome).

In this context, "verifiable" means an event is observable not only by the parties to the contract, but also by any third party (*e.g.,* a judge) who might be called upon to adjudicate a dispute. The focus on verifiable events is motivated as follows. Were an outcome contingent on an unverifiable event (*i.e.,* one not observable to the third party), then there would be no way for the third party to judge the extent of breach of contract (if any) or even who breached (if anyone did). Hence, a contractual obligation that is contingent on an unverifiable event cannot be effectively enforced by a third party.

In the incomplete-contracts literature, it is standard to assume that the parties to a contract can observe events that cannot be verified by any judge. In the parlance of the literature, such events are described as observable, but not verifiable. As we will see, the parties could ideally wish to base their contract on such observable, but unverifiable events.

Enforcement would also be difficult if one of the parties to the contract couldn't observe an event on which an outcome was contingent (for example, as in the case of a consumer who cannot tell whether a mechanic has properly repaired an automobile). That party would not know the extent to which the other party was out of compliance with the contract (if he was). Such ignorance would, thus, make a contract impossible to enforce even through private sanctions of the sort discussed in §5.4.2 and §5.4.3 below. For this reason *observability* is considered a minimal informational requirement for an event to define a contractual contingency.

To be more formal, if we take Ω to be the set of verifiable events (with ω a representative element) and $\mathcal{A}$ to be the set of outcomes, then a contract can be seen as a mapping, $C : \Omega_C \to \mathcal{A}$, where $\Omega_C \subseteq \Omega$. Contractual *in*completeness can, then, be seen as situation in which the parties to the contract would or should ideally wish to base their contract on some set other than Ω_C (for example, Ω if $\Omega_C \subset \Omega$; or the set $\mathcal{O}$ of observable events).

Within the economics literature, the terms verifiable and observable are typically employed without any explicit consideration of the costs associated with investigation, measurement, documentation, or monitoring; that is, activities necessary to make information useful contractually. To an extent, one implicit set of assumptions typically employed is that observable events are observable at no cost, similarly for verifiable events, while *un*verifiable events would be verifiable only at a prohibitive cost (possibly infinite). As Hermalin and Katz (1991) point out, events can be "partially" verifiable in the sense that although a third party cannot observe them with the precision of the parties to the contract, a third party can still observe some information about the event that causes him or her to update his or her beliefs about what happened in a way useful for adjudication (see discussion below in §4.1.3 and §4.3.1). In any case, whatever the costs of observation and verification, they are almost always taken to be exogenous to the model. Endogenizing these costs through the design of measurement or monitoring systems remains, for the most part, an important area for future research.

4.1.1 Literal incompleteness and unmapped contingencies

Following Hermalin and Katz (1993), we make a distinction between literal incompleteness, which we consider now, and economic incompleteness, which we address in §4.1.3.[49]

A contract is literally incomplete if an event or contingency can arise that is not anticipated by the contract; hence, the contract is silent with respect to what should happen given this event or contingency. Literal incompleteness corresponds to a situation in which there are elements of Ω not in Ω_C. Ayres and Gertner (1989) refer to such as elements as "gaps." Other scholars have referred to them as *unforeseen contingencies*. Let $\bar{\Omega}_C$ denote the set of gaps or unforeseen contingencies (*i.e.,* $\bar{\Omega}_C = \Omega - \Omega_C$).

The assumption of literally incomplete contracts has played an important role in law & economics (both implicitly, as in Shavell, 1980, and Rogerson, 1984, and explicitly, as in Goetz and Scott, 1981, Ayres and Gertner, 1989, and Hadfield, 1994). Nonetheless, as Hermalin and Katz (1993) observe, it is a potentially problematic assumption, because it is so easy to complete contracts by adding a stereotypical residual ("none-of-the-above") clause to a contract. That is, literal *completeness* can be achieved simply by adding a clause that states, "if an event (contingency) other than those listed above occurs, then the outcome shall be . . . "

In order to explain literally incomplete contracts, consequently, it is not enough to assume that the parties fail to foresee some contingencies. It would seem necessary to assume also that the parties fail to foresee that they could fail to have foreseen some contingencies.

Alternatively, one might assume the gap is rational, insofar as there is some affirmative reason why the parties deliberately left contingencies unmapped. One reason could be that the parties are content to let the courts apply a particular outcome, and this outcome is no worse than what the parties might have provided on their own. Under traditional common law, for instance, courts often responded to contractual gaps by treating the contract as a nullity, on the grounds that there had been no "meeting of the minds." The result would be that the parties would be left where they lay when the contingency arose (although parties who had partially or fully performed might be entitled to a refund or to restitution of any value conferred on the counterparty). Modern-day courts, in contrast, are more likely to react to a gap by trying to fill it with either an objectively reasonable term, or with their best guess as to what the parties would have wished had they negotiated over the contingency in question.

[49] Ayres and Gertner (1992) refer to the first type of incompleteness as "obligational" and second as "insufficiently state contingent."

(Although if the gaps are too significant or the parties' hypothetical wishes too unclear, the traditional approach of declaring the contract to be at an end is still a real possibility.)

To the extent that courts apply a predictable default rule in such situations (or even an unpredictable one), one could say that there is no such thing as a literally incomplete contract—rather, the implicit residual clause is just that the parties agree to go to court—or follow existing legal requirements that they pursue private arbitration or other method of dispute resolution—and abide by any result that is reached there. At a semantic level, such logic is unassailable, if perhaps unappealing to those worried about truly naïve parties who indeed failed to foresee that might have failed to have foreseen a contingency.

4.1.2 Linguistic under- and overdetermination

In many situations in which a contract does not point to a clear outcome, the problem is not that the parties have said nothing; rather, it is that they have said too little or too much. For example, the parties might provide multiple and inconsistent provisions dealing with the same event. Alternatively, the parties might use terms that admit multiple meanings or that depend on other terms of the contract.

One could treat such cases as equivalent to contractual gaps, arguing that, if the parties have not settled a term definitively, they have not settled it at all, but this interpretation does not appear to comport with the behavior of either legal institutions or actual bargainers. An alternate approach, accordingly, would be to model contracts not as single-valued functions, but as multi-valued correspondences. Formally, instead of representing a contract as a mapping of the form $C : \Omega_C \to \mathcal{A}$, we could represent it as a mapping of the form $C : \Omega_C \to P(\mathcal{A})$ where $P(\mathcal{A})$ denotes the power set of $\mathcal{A}$.[50] Under this interpretation, incompleteness would arise whenever the contract mapped an event to a set with more than one outcome, so that the person applying the contract would have to choose among those outcomes based on extra-contractual factors.

This type of incompleteness has received less attention in the literature, but some recent work has begun to explore its implications. For instance, a recent paper by Hart and Moore (2004) models a contract as a list of outcomes to which the parties are restricted. The determination of which outcome from that list is implemented occurs through *ex post* bargaining. Through this approach, they show that a relatively loose contract (*i.e.,* a relatively long list) preserves flexibility, allowing the parties to make use of new information that arises, but at the expense of distorting *ex ante* investments due to the increased danger of holdup. Similarly, Ben-Shahar (2004) has argued that courts should respond to such incompleteness by granting both parties the option to enforce the agreement on whatever terms are most favorable to the counterparty; such a response would preserve the benefits of whatever bargain the parties had already reached, while allowing them to enjoy further gains from trade through

[50] A sigma field should $\mathcal{A}$ be non-denumerable.

further negotiation.

A criticism of Hart and Moore (2004) is that relies heavily on some strong assumptions. First, they rely heavily on the observable but unverifiable distinction. As Hermalin and Katz (1991) and Maskin and Tirole (1999) note, there are reasons to doubt this distinction in many contexts. More critically, they also rely on the ability of parties to carry out what would normally be seen as incredible threats. Hart and Moore justify these by appealing to the psychological tendency of people to sacrifice resources in order to get revenge against those who have, in their eyes, mistreated them (see Rabin, 1993, for a discussion of such psychological effects in economic contexts). While such tendencies might exist among individuals, one is hesitant to assume that sophisticated parties (*e.g.,* firms) would be prey to such base emotions.

4.1.3 Economic incompleteness and coarse mappings

Much of the *economic* literature on incomplete contracts has focused on a different standard of incompleteness: A contract is incomplete when the set of verifiable events is not the same as the set of observable events.[51] That is, even if the contract is literally complete (*i.e.,* $\bar{\Omega}_C = \emptyset$) it would be judged *economically* incomplete if $\Omega \neq \mathcal{O}$, where $\mathcal{O}$ is the set of observable events. In this case, the parties would benefit from being able to condition their contractual obligations more finely (for instance, by allowing an excuse when performance is commercially impractical), but they cannot do so because the condition cannot be effectively enforced (for instance, because the court cannot tell the difference between commercial impracticability and mere seller recalcitrance).

As noted earlier, it is natural to assume $\mathcal{O}$ is "larger" than Ω. In the literature, there are essentially two ways this assumption is modeled.[52] One is to define Υ as the set of observable, but unverifiable events (with representative element υ) and take

$$\mathcal{O} = \Upsilon \times \Omega\,.$$

In other words, an observable event is a vector consisting of events that the parties can observe, but not verify, and events that the parties can both observe and verify.[53] Under this formulation, a verifiable event ω reveals not only that

[51] Obviously, both sets must be defined in some way over *relevant* events; that is, we don't want to say the sets differ if the observable, but unverifiable events are irrelevant to the contracting situation. Defining relevance is, however, difficult insofar as one strain of the literature (*e.g.,* Hermalin and Katz, 1991; Maskin and Tirole, 1999; Edlin and Hermalin, 2001) has been devoted to showing when the observable, but unverifiable distinction is unimportant; that is, situations in which the parties can achieve the same outcomes with incomplete contracts as they could with complete contracts. In such situations, it would be misleading to say that the observable and verifiable sets are the same over relevant events, because equilibrium play, but not outcomes, would be different were it feasible to base contracts on the set of observable events. A definition of relevant events must, therefore, account for potential differences in the ensuing game, as well as outcomes.

[52] An example of the first way is Grossman and Hart (1986). An example of the second way is Anderlini and Felli (1994).

[53] It is worth noting that both Υ and Ω could, themselves, be vector spaces.

ω has occurred, but also that

$$v \in \mathcal{O}_\omega \equiv \{v | (v, \omega) \in \mathcal{O}\}$$

(the set $\mathcal{O}_\omega$ is called the *section of* $\mathcal{O}$ *at* ω). Because, absent additional assumptions, $\mathcal{O}_\omega = \Upsilon$ for all $\omega \in \Omega$, it might at first seem that knowing that v is in $\mathcal{O}_\omega$ is not useful information. But if some (v, ω) pairs are impossible, then some sections satisfy $\mathcal{O}_\omega \subset \Upsilon$ and, therefore, learning ω can lead to inferences about which v occurred.

In such situations, a so-called "forcing" contract (as in Harris and Raviv, 1978) could be useful. Suppose the parties wish to induce one party, "the actor," to choose a specific $\hat{v}$. If there are ω such that $\hat{v} \notin \mathcal{O}_\omega$, then those ω constitute proof that the actor has not chosen the desired action and the contract can, therefore, threaten the actor with sufficient punishment should such ω occur that he would never choose an $v \in \mathcal{O}_\omega$. Hence, some undesirable actions can be avoided. Ideally, if there is a subset $\Omega_0 \subset \Omega$ such that, for all $v \neq \hat{v}$, $v \in \mathcal{O}_\omega$ for some $\omega \in \Omega_0$, but $\hat{v} \notin \mathcal{O}_\omega$ for any $\omega \in \Omega_0$, then $\hat{v}$ can be achieved by a contract that sufficiently punishes the actor for any $\omega \in \Omega_0$ and rewards him only for $\omega \in \Omega \backslash \Omega_0$.[54]

Even if all (v, ω) pairs are possible—so $\mathcal{O}_\omega = \Upsilon$—the *conditional distribution* of v given $v \in \mathcal{O}_\omega$ may differ from the distribution given $v \in \mathcal{O}_{\omega'}$, $\omega \neq \omega'$. Such differences in distributions can be very powerful, often allowing contractual solutions that replicate the outcomes that would have occurred were the parties able to contract on $\mathcal{O}$ directly (see discussion in §4.3.1; also Hermalin and Katz, 1991).

The second way economic incompleteness can be understood is to take Ω to be a *partition* of $\mathcal{O}$; that is, each ω is a subset of $\mathcal{O}$, any given element of $\mathcal{O}$ can be in only one ω, and the ωs, as a class, contain all the elements in $\mathcal{O}$.[55] Of course, to make this interesting, there must be at least one ω with two or more elements of $\mathcal{O}$ in it; that is, for this ω at least, learning ω does not perfectly reveal the observable information. Conversely, economic *completeness* would correspond to a situation in which each ω contained a single element of $\mathcal{O}$.

The standard interpretation of this second representation of economic incompleteness is that there is inherently no difference between what is observable and what is verifiable, except that it is impossible or prohibitively expensive to describe the observable events with sufficient precision in a contract.[56] Instead, the observable events can only be described coarsely; so that different observable events get lumped together in a single ω. For example, consider the "contract" between us, the authors of this chapter, and the editors of this volume. Without effectively writing the chapter themselves, it is impossible for the editors to stipulate fully what the chapter should look like; the best they can do is stipulate the topics to be covered and the overall length of the chapter.

[54] The notation $\mathcal{S} \backslash \mathcal{T}$ denotes the set whose elements are in $\mathcal{S}$ but not $\mathcal{T}$ (*i.e.*, $\mathcal{S} \backslash \mathcal{T} = \mathcal{S} \cap \mathcal{T}^c$).

[55] Formally, the ωs satisfy

$$\omega \cap \omega' = \emptyset, \ \forall \omega \neq \omega' \text{; and } \bigcup_{\omega \in \Omega} \omega = \mathcal{O}\,.$$

This second way of representing economic incompleteness can be seen as a special case of the first in which $\mathcal{O} \subset \Upsilon \times \Omega$ such that $\mathcal{O}_\omega \cap \mathcal{O}_{\omega'} = \emptyset$ for any two ω, ω' and $\bigcup_{\omega \in \Omega} \mathcal{O}_\omega = \Upsilon$.

[56] Typically these costs are taken to be "writing costs." They could also, as Rasmusen (2001) notes, be due to "reading" costs.

4.1.4 Discussion

Each of these models of contractual incompleteness has some appeal, though none is entirely satisfactory as a complete account of the phenomenon. To illustrate, consider the case of *Spaulding v. Morse*, in which a divorcing couple entered into a maintenance agreement under which the husband agreed to pay child support of $1200 per year until the couple's son entered some "college, university, or higher institution of learning ...," and $2200 per year for four years thereafter. The dispute arose when the son completed high school and was immediately drafted into the US Army, and the husband took the position that his contractual obligation was suspended while the son remained in military service. One can view this dispute as arising from literal incompleteness if we say that the contract simply did not provide for the case in which the son was drafted into the army. In order to take this view, however, we must read the term in question in a non-literal way, since a literal reading makes it appear to be a residual clause. (That is, it appears to say that so long as the son never started college—*e.g.,* if he chose to pursue a military career or even if he were killed in combat—the husband would be required to make yearly payments in perpetuity.)[57]

One can alternatively view the dispute as arising from linguistic underdetermination if we focus on the phrase "child support." Does the phrase simply refer to any payments made for the benefit of the son, or is it instead limited to payments that are objectively necessary for the son's adequate maintenance? Ordinary language is probably not precise enough to provide a definitive answer to this question, and the best response may be to say that the phrase can mean either or both.

Finally, we could view this dispute as a case of economic incompleteness if we assume that what the parties actually wished to do was to condition payment on the son's need for maintenance (an observable but perhaps not verifiable event) but that they instead conditioned on the son's not yet having completed four years of college (a coarser but perhaps more easily verifiable proxy). Had the parties been able to describe the former condition at reasonable cost in their divorce settlement, the whole dispute could have been avoided. On the other hand, it is not clear that the couple had such a clear purpose in arriving at the clause in question. Another possibility is that the husband simply wished to

[57] Nobody in the case made this argument, however, not even the wife. Instead, the court found in favor of the husband, on the grounds that the unstated purpose of the clause was to provide financial support for the son in his minor years, and once the Army had taken responsibility in this regard, the purpose had been served.

pay as little as possible, the wife wished to receive as much as possible, and the clause was a rough compromise.

As the example indicates, modeling contracts as a formal mapping from events to outcomes is useful, but it could fail to capture two important real-world features of the situation—namely, that parties do not always possess a well-specified understanding of the domain of possible events, either *ex ante* or *ex post* and that they write contracts in ordinary language with all its ambiguities, rather than in the formal but restricted language of mathematics. This is not to say, however, that such failures necessarily invalidate or diminish the insights of formal modeling. Nor is it to say that formal modeling can't deal with these features should they be important. Certainly, the first could be readily modeled by incorporating some uncertainty over what the domain of events might be. The second is potentially trickier, but techniques such as assuming asymmetric information with regard to the meaning of the terms, uncertainty over their interpretation, or even suspension of the usual common-priors assumption could be useful. With all that in mind, the next subsection discusses possible economic causes of contractual incompleteness.

4.2 The sources of contractual incompleteness

4.2.1 Bounded rationality

A common explanation for incomplete contracts, especially literally incomplete contracts, is bounded rationality.

The simplest "model" of bounded rationality is that people make mistakes. They fail to foresee all possible contingencies and, thus, their contracts suffer from unforeseen contingencies. However, as noted above (§4.1.1), failure to foresee certain contingencies need not generate incomplete contracts unless the parties also fail to foresee that they may fail to foresee certain contingencies. If the parties recognize their imperfect foresight, then they can complete any contract with a residual ("none-of-the-above") clause.

Of course, the fact that the parties *could* complete any contract with a none-of-the-above clause does not mean that they would wish to do so. After all, they may fear that the ideal response to different unforeseen contingencies varies with those contingencies. By its nature, a none-of-the-above clause is a one-size-fits-all solution. Hence, the parties may wish for flexibility should an unforeseen contingency occur; legal proceedings can provide such flexibility insofar as the court's ruling will typically depend on the contingencies that have occurred. How the courts should respond to these "intentional" gaps has become an issue in the literature (see, *e.g.,* Ayres and Gertner, 1989, 1992). We discuss this literature in §4.4 below.

To an extent, deciding how the courts should respond depends on the reason for unforeseen contingencies in the first place. As, however, E. Posner (2003a) points out, economic theory does not offer particularly good or widely accepted means of modeling unforeseen contingencies (or bounded rationality more generally).[58] To the extent models of unforeseen contingencies exist, they are too abstract to be readily utilized in the study of contracts. Admittedly, one could simply appeal to the existence of gaps in actual contracts to justify assuming literal incompleteness. Such an atheoretic rationale, however, makes it difficult to analyze important issues such as how the courts would know whether gaps are truly mistakes or have arisen because one side or the other is being strategic. This, in turn, affects how the courts should respond to gaps (see Ayres and Gertner, 1989, 1992, and Shavell, in press).

Although some parts of the literature appeal to unforeseen contingencies as a possible motivation for *economic* incompleteness (see, *e.g.,* Maskin and Tirole, 1999), it is difficult to reconcile such bounded rationality with the "dynamic-programming" rationality (to use Maskin and Tirole's term) that the incomplete-contracts literature typically assumes. Other, more rational, explanations for economic incompleteness strike us as more plausible. We explore some of these explanations in the next few sections.

4.2.2 Describability and contracting costs

As we observed, a complete contract between us and the book's editors about what this chapter should be is rather impractical. If the editors spelled out in great detail what they wanted, they would, in effect, be writing the chapter themselves. Their costs of doing presumably outweigh whatever benefits such detailed contracting would generate. More generally, there is a cost to writing fine-tuned contracts. In some instances, such as our chapter example, it is simply impractical to describe the relevant contingencies finely enough.

Describability. Because describing something can be viewed as an algorithm, it follows from the mathematics of computability (see, *e.g.,* Boolos et al., 2002) that it is conceivable that some events or contingencies cannot be described, in the sense that no algorithm for describing them would ever finish. This point is made by Anderlini and Felli (1994). An intuitive sense of this idea can be had by recalling the well-known math fact that the constant π cannot be given a decimal representation; no algorithm could ever complete the task of fully describing π as a decimal number.

Indescribability *per se*, however, seems a doubtful explanation for economically incomplete contracts for at least two reasons. First, the parties might be able to approximate their desired description arbitrarily well (*e.g.,* 3.1415 might be a good enough approximation of π); therefore, they suffer arbitrarily little from the indescribability of a contingency. In fact, Anderlini and Felli prove just such an approximation theorem. Second, the parties to the contract need to have internal descriptions of the relevant events or contingencies to know if they have occurred; hence, if they cannot write descriptions, it is hard to under-

[58] For a general survey of recent modeling on bounded rationality see Rubinstein (1998). For a more specific survey on unforeseen contingencies see Dekel et al. (1998) (also see Dekel et al., 2001).

stand how they know the relevant descriptions. In other words, if we take the feasible contractual contingencies to be subsets of some partition of the relevant state space, then why isn't the set of observable contingencies this same partition?[59] And, if it is, then there is no practical difference between observable and verifiable; that is, the contract is *not* economically incomplete.

Description costs. A better way to proceed is to assume that the relevant states can be described (or described arbitrarily finely), but to recognize that the costs of doing so can sometimes be so large as to render detailed descriptions impractical (see, *e.g.,* Dye, 1985). The chapter-writing example given earlier is an example of such a situation.

Having recognized that contract costs are increasing in the details of the contract,[60] the next question is how to balance the marginal gain from more details against these marginal costs? As it turns out, however, in many settings the marginal *benefit* of adding details is zero; as Hermalin and Katz (1991) and Maskin and Tirole (1999) show, even rough descriptions can be sufficient for the parties to do as well as they could were it practical to write very detailed contracts. We return to this point later.

Another reason the marginal benefits to the parties could be small is that the courts will fill in or interpret the missing or vague terms (see, *e.g.,* Ayres and Gertner, 1989, 1992, or Shavell, in press). Hence a model of how the courts enforce incomplete contracts is essential to an analysis of the effort that parties should make to make their contracts more precise.

Evaluation and expertise. It is possible that even if a judge, or other adjudicator of contractual dispute, can observe a relevant variable, he or she may not properly evaluate it. His or her evaluation could differ from that of the parties to the contract, perhaps because of a lack of expertise in the relevant field. This idea can be captured by assuming that, when the parties observe x, the judge observes (concludes) $x+\varepsilon$, where ε is the judge's error in evaluation. The error, ε, could also be introduced because the judge reads or interprets the contract differently than the parties intended. (The model, which we consider in §4.4.1 *infra*, on form versus substance also touches on this point.)

While such errors might, at first, seem to render the relevant variable essentially unverifiable, the parties can, in some circumstances, do as well as they would were the judge to observe the relevant variable without error (Hermalin and Katz, 1991). We return to this point later in §4.3.1.

[59] Let Σ be the state space. What can be described, the verifiable events, are subsets of Σ (*i.e.,* $\omega \subset \Sigma$ for all $\omega \in \Omega$, where $\bigcup \omega = \Sigma$ and $\omega \cap \omega' = \emptyset$). But if, as seems reasonable, what the parties can describe in writing is the same as what they can describe to themselves, then each ω is, then, an element of $\mathcal{O}$ and $\mathcal{O} = \Omega$. [In a sense, this is reminiscent of Wittgenstein's (1922) Proposition 7 that what we can't think about, we can't verbalize.]

[60] One can model contracts being more detailed by assuming that there is an order, $\succ$, on partitions of $\mathcal{O}$, such that $\Omega_i \succ \Omega_j$ (*i.e.,* a contract based on Ω_i is more detailed than one based on Ω_j) if for each $\omega_i \in \Omega_i$ there exists an $\omega_j \in \Omega_j$ such that $\omega_i \subseteq \omega_j$ and for at least one $\omega_j \in \Omega_j$ there exist ω_i and ω_i' in Ω_i such that $\omega_i \cup \omega_i' \subseteq \omega_j$.

4.2.3 Complex environments

Segal (1999a) suggests that economically incomplete contracts can arise when the contracting environment is complicated. In his model, the parties wish to trade one "widget" from a set of different widgets. As the number of potential contingencies (*i.e.,* widgets in the set) rises, the optimal second-best contract does increasingly worse. In the limit, the optimal second-best contract does no better than the simple incomplete contract (equivalently, in his model, no contract). The reason for this worsening performance is that, as the number of contingencies rises, the number of incentive constraints increases. As in all optimization problems, as the number of constraints increases, the optimality of the solution decreases (at least weakly). Given that writing complex contracts is costly, there is a cutoff point at which the environment is so complex that the simple, less costly contract is superior to a complex contract.

Segal's model, while elegant, rests on a number of strong assumptions. Among these is the assumption that almost no variable of interest is verifiable, despite many of them being observable (his Assumption 2, page 60). If either the optimal widget to trade is known *ex ante* or which is the optimal widget to trade can be *verified ex post*, the model collapses. As we have observed, the assumption of observable but unverifiable can be a questionable one, especially in many of the contexts to which this model is intended to apply.

4.2.4 Asymmetric information

As noted in §2.3.2, informational asymmetry at contracting can lead to distortions in the contract that is written. Spier (1992) builds on this idea by suggesting that one consequent distortion could be contractual incompleteness.

A simple model can serve to illustrate her idea. Suppose a manufacturer needs a part for one of its manufacturing machines. The manufacturer can have the part sent by a default carrier, in which case the part will certainly arrive in two days. Normalize the default carrier's price to be 0. Alternatively, the manufacturer can contract with an express carrier. If the express carrier makes no special efforts, the part will arrive in one day with probability $1-q \in (0,1)$; with probability q it arrives the day after that. For simplicity, take the express carrier's cost when it makes no special efforts to be 0.[61] Alternatively, by spending $c > 0$, the express carrier can ensure the part arrives in one day. Let $L > 0$ be the amount per day that the manufacturer loses from the idle machine. If the manufacturer selects the default carrier, its payoff is $-L$. If it selects the express carrier, but no special efforts are made, its expected gross payoff is $-qL$. If it selects the express carrier and special efforts are made, its expected gross payoff is 0; but, in this case, the express carrier's gross payoff is $-c$. Assume that there are two types of manufacturer, 0 and 1, with

$$qL_1 > c > qL_0\,.$$

[61] Making it a positive amount does not alter the analysis materially, while setting it to zero simplifies the notation.

Observe that it would be welfare maximizing to have the express carrier make special efforts for a high-value manufacturer (type 1), but not for a low-value manufacturer (type 0). Let $\theta \in (0,1)$ be the probability that the manufacturer is type 1. Hence,

$$\mathbb{E}L = \theta L_1 + (1-\theta)L_0 \,.$$

Assume that the manufacturer's type is its private information.

To close the model, we need to make some assumptions about bargaining between the manufacturer and the express carrier. Assume that if the express carrier can't infer the manufacturer's type, then the express carrier and manufacturer set the price through Nash (1950) bargaining over the expected surplus, $\mathbb{E}L - q\mathbb{E}L$. In that case, the price is

$$p_u = \frac{1}{2}(1-q)\mathbb{E}L \,.$$

Assume the parameters are such that

$$-qL_0 - p_u > -L_0$$

(this clearly holds for low values of θ). Because it is welfare reducing for a low-value manufacturer to request special efforts, we can assume that the express carrier interprets any request for special efforts to be from a high-value manufacturer. In this case, total surplus is $L_1 - c$. If bargaining is again Nash bargaining, then the resulting price is

$$p_i = \frac{L_1 + c}{2} \,.$$

Finally, consider a high-value manufacturer who is deciding between asking for special efforts—and thus revealing its identity—or pooling with low-value manufacturers and not asking. Its expected net surplus in the first case is $-p_i$. Its expected net surplus in the second case is $-qL_1 - p_u$. It is readily shown that there exist parameter values such that latter exceeds the former (*e.g.,* $c = 2$, $L_1 = 5$, $L_0 = 3$, $q = 1/2$, and $\theta = 1/4$ would work). Hence, it is possible that asymmetric information leads to a form of contractual incompleteness: Although it would be welfare improving for a high-value manufacturer and the express carrier to have a contingency in their contract requiring the carrier to make special efforts, they don't have such a contingency in equilibrium.

4.2.5 Verification costs

In much of the contracting literature, it is assumed that, if information is verifiable, it can be verified costlessly. In real life, however, information is made verifiable at a cost (*e.g.,* surveillance monitoring, auditing, and record keeping). If such costs are large relative to the benefits that complete contracting affords, the consequence will be incomplete contracts.

There is a strand of the literature on "costly state verification," starting with Townsend's (1979) seminal article, that considers, in part, the consequences of an inability to verify payoff-relevant variables due to the cost of verification. The principal concern of this literature has been on financial contracting; particularly the decision to use debt rather than equity.[62]

4.2.6 Dynamic inconsistency with respect to renegotiation

A common, although not universal, view in the economics literature is that parties never "leave money on the table" even on out-of-equilibrium paths. More precisely, should the parties ever reach a point at which it is common knowledge that the allocation dictated by the contract is Pareto inferior, the parties will renegotiate the contract. This ability to renegotiate undermines the value of contracts that rely on Pareto-inferior allocations as threats should one or both parties deviate from their contractual obligations. In the lingo of the literature, such contracts fail to be *renegotiation proof*. That contracts be renegotiation proof is, therefore, often a requirement imposed in the literature.

To illustrate the concept of renegotiation-proofness and how it can lead to incomplete contracts, we briefly review the Fudenberg and Tirole (1990) model of renegotiation in agency. In a standard agency model, to induce the agent to work hard, the principal and agent agree to a contract that makes the agent's compensation contingent on the outcome (*e.g.,* salespeople's compensation is frequently tied to their sales). Such a contract is, however, generally not first-best optimal because it causes a risk-averse agent to bear risk. Fudenberg and Tirole consider what happens in the standard model if there is a period after the agent has committed his effort but before its outcome is known (*e.g.,* after he is back from his sales calls, but before the orders arrive) during which the principal and agent can renegotiate. Observe that if, as is typically assumed, the principal is risk neutral, then it is common knowledge that leaving the agent with risky compensation is inefficient; the parties should renegotiate to a fixed, non-contingent wage for the agent. However, a rational agent would forecast such renegotiation and, thus, that his compensation will ultimately be unrelated to his effort. But if it is unrelated to his effort, then he has no incentives to work hard. While Fudenberg and Tirole show that there are ways to restore *some* incentives, the potential for renegotiation limits their effectiveness. Moreover, a consequence could well be that the parties forgo using an incentive contract even though they would use one were renegotiation not possible.

This last point shows how the threat of renegotiation can lead the parties to use less complete contracts than they would were renegotiation infeasible. That is, while they might wish to write a contract contingent on the outcome, they can't. Schwartz and Watson (2003) consider the relation between contractual complexity and the feasibility of renegotiation in greater depth, with a particular emphasis on law & economics issues.

[62] See, for instance, Webb (1992) or Hart and Moore (1998).

4.3 Consequences of contractual incompleteness

4.3.1 Does incompleteness matter?

A debate has emerged in economics as to whether incomplete contracts matter. Specifically, even if it is accepted that contracts are incomplete, are there ways for the parties to effectively contract around the incompleteness?

Hermalin and Katz (1991) offer one model in which incompleteness doesn't matter. Let Υ denote the set of observable, but unverifiable variables. Assume that Υ is finite. Let Ω denote the set of verifiable variables. Take it to be finite as well, with N elements indexed by n. Let there be two parties to the contract: the agent and the principal.

The agent chooses $v \in \Upsilon$. The principal has preferences over the vs. Suppose that the agent is willing to sign a contract with the principal if his equilibrium (expected) utility is at least some reservation level, which we can normalize to be zero. Let w denote the agent's gross utility and let $w - d(v)$ denote his net utility ($d : \Upsilon \to \mathbb{R}_+$). Assume it costs the principal $c(w)$ if the agent receives gross utility w, where $c(\cdot)$ is increasing and strictly convex.[63] Finally, assume that the conditional distribution of the verifiable variable, ω, depends on the v chosen; specifically, let $\pi_n(v) = \Pr\{\omega_n | v\}$ and let

$$\boldsymbol{\pi}(v) \equiv \big(\pi_1(v), \ldots, \pi_N(v)\big)$$

be the density over Ω induced by v.

Suppose that the principal wants the agent to choose a specific v^*. If we assume that the principal has all the bargaining power, her ideal would be a contract in which the agent agrees to choose v^* in exchange for a net utility of 0, which costs the principal $c\big(d(v^*)\big)$. Except in the special case where v^* minimizes $d(v)$, this ideal contract is infeasible because v is not verifiable and the agent prefers an v other than v^*.

A feasible alternative is a contingent-compensation contract; the principal agrees to provide utility level w_n should verifiable outcome ω_n occur. Let $\mathbf{w} = (w_1, \ldots, w_N)$. Such a contingent contract will be agreeable to the agent and induce him to select v^* if

$$\boldsymbol{\pi}(v^*) \cdot \mathbf{w} - d(v^*) \geq 0 \text{ and} \tag{7}$$

$$\boldsymbol{\pi}(v^*) \cdot \mathbf{w} - d(v^*) \geq \boldsymbol{\pi}(v) \cdot \mathbf{w} - d(v) \text{ for all } v \in \Upsilon \backslash \{v^*\}. \tag{8}$$

Condition (7), the individual rationality constraint, is the condition that the agent agree to the contract. Condition (8), the set of incentive compatibility constraints, is the condition that the agent prefer choosing v^* to any other v.

Suppose the principal has all the bargaining power. It is a well-known result (see, *e.g.,* Grossman and Hart, 1983) that if a contract (a vector $\mathbf{w}$) exists satisfying expressions (7–8), then there exists one such that (7) holds as an equality. Because that contract minimizes the principal's expected cost, she will offer that contract. Under general conditions (see, *e.g.,* Grossman and Hart, 1983), $\text{Var}(w|v^*) > 0$. Hence, by Jensen's inequality, $\mathbb{E}\{c(w)|v^*\} > c\big(d(v^*)\big)$; that is, even if the principal can devise a contract to implement v^*, it will cost her more than the ideal contract would were v verifiable. It would seem, therefore, that there is an adverse consequence to v not being verifiable.

This last conclusion, however, need not follow if the parties have time between the realizations of v and ω to renegotiate the contract. Suppose that the principal retains all the bargaining power in renegotiation.[64] Because, here, the principal observes v, the problem identified by Fudenberg and Tirole (1990) is not triggered by renegotiation (see discussion in §4.2.6). Because the principal's cost increases with the variability of w, the principal would offer, in renegotiation, a non-contingent w, $\bar{w}$. Because the agent's cost from selecting v is sunk at the point of renegotiation, he will accept $\bar{w}$ if and only if $\bar{w} \geq \boldsymbol{\pi}(v) \cdot \mathbf{w}$. As she has all the bargaining power, the principal will choose the lowest such $\bar{w}$, namely the one that just equals $\boldsymbol{\pi}(v) \cdot \mathbf{w}$. Hence $\bar{w}$ is a function of v, which we indicate by writing $\bar{w}(v)$. By Jensen's inequality, $c\big(\bar{w}(v)\big) \leq \mathbb{E}\{c(w)|v\}$; that is, renegotiation lowers the principal's cost. Moreover, because $\bar{w}(v) = \boldsymbol{\pi}(v) \cdot \mathbf{w}$, it is readily seen that anticipation of such renegotiation does *not* change the individual rationality nor incentive compatibility constraints (7–8). Moreover, from (7), we see that $\bar{w}(v^*) = d(v^*)$. We therefore have

Proposition 6 *Suppose that renegotiation is possible. If a contract* $\mathbf{w}$ *exists that satisfies individual rationality and incentive compatibility (i.e., expressions 7 & 8), then even though* v *is* un*verifiable, the principal can achieve the same outcome as she could if it were verifiable.*

In other words, in this context, the distinction between observable and verifiable is not relevant to the outcome.

Proposition 6 rests on the existence of a $\mathbf{w}$ satisfying expressions (7–8). Fortunately, the conditions for such a $\mathbf{w}$ to exist are rather minimal. Somewhat loosely, it is sufficient that the distribution over ω induced by v^* be distinct from the distributions induced by the other v.[65]

Proposition 6 also rests on the supposition that renegotiation is possible; that is, that there be a lag between when v is chosen and when ω is realized. Renegotiation would seem possible in many contexts of interest. For instance, if ω is a judge's observation and, thus, a ruling as to the value of v, then the parties presumably have time to renegotiate prior to the judge's verdict.

A provocative way to summarize this is as follows: Even when critical variables are not verifiable, parties can often present some relevant evidence that

[63] A natural interpretation is that the agent is risk averse, while the principal is risk neutral, in which case $c(\cdot)$ is the inverse function of the agent's utility function.

[64] Edlin and Hermalin (2001) extend the analysis to a broader set of bargaining games in renegotiation.

[65] Formally, it is sufficient that $\boldsymbol{\pi}(v^*)$ not be an element of the convex hull of $\{\boldsymbol{\pi}(v)|v \in \Upsilon\backslash\{v^*\}\}$. As Hermalin and Katz (1991) discuss, this condition is readily met in most situations of interest. As Edlin and Hermalin (2001) show, for more general bargaining games in renegotiation, a slightly stronger condition could be required: there exist a subset Ω^* of Ω such that $\Pr\{\omega \in \Omega^*|v^*\} > \Pr\{\omega \in \Omega^*|v\}$ for all $v \in \Upsilon\backslash\{v^*\}$.

is informative, in a noisy way, about the critical variables. If renegotiation is feasible, then, by contracting on this relevant evidence, the parties can typically write contracts *as if* the critical variables were verifiable.

Renegotiation *à la* Hermalin and Katz is a particular type of game or mechanism. If a larger set of mechanisms is considered, one would find that the observability-verifiability distinction is irrelevant under an even wider set of circumstances. In a series of articles, Maskin and Tirole do just that, considering mechanisms more generally (see, *e.g.,* Maskin and Tirole, 1999; see, also, Tirole, 1999, for an overview).

In a mechanism, the parties send messages to the mechanism arbitrator (possibly the judge or other dispute adjudicator) about the variables they can observe. As Farrell (1987b) notes, perhaps the earliest recorded use of a "mechanism" is the one King Solomon used to determine which of two women claiming to be the mother of a child was the true mother.

A review of mechanism design is beyond the scope of this chapter. We can, however, give a sense of its application in this context. Suppose that there is some task that either of two parties, 1 or 2, could do. *After* contracting, the parties observe their costs $(c_1, c_2) = v$ of doing the task, but cannot verify them. Suppose that $c_i \in \{L, H\}$, where L (low) is smaller than H (high).

Efficiency requires that the party with the lower cost do the task. The parties might further wish to have the one who doesn't do the task partially compensate the other (*i.e.,* pay some portion of the cost). But, by assumption, the parties cannot write such a contract directly because it would be contingent on the realizations of their costs. Fortunately, however, a mechanism can be constructed that replicates the desired outcome: Both parties simultaneously announce their individual costs (*i.e.,* 1 announces c_1 and 2 announces c_2). The following rules govern what happens next:

- If they announce the same cost, then a coin is flipped. The loser of the coin flip has to do the task, but is paid $\hat{c}/2$ by the winner, where $\hat{c}$ is the commonly announced cost.
- If they announce different costs, then the lower-cost party must do the task, but she is paid $H/2$ by the high-cost party.

If the parties announce truthfully, then the allocation of the task will be efficient. It remains, however, to check that the parties will announce truthfully in equilibrium (*i.e.,* that the mechanism is incentive compatible). There are three cases to consider. In each, we are assessing whether truth telling is a best response to truth telling; and, therefore, does a truth-telling equilibrium exist.

1. $v = (L, L)$. If a party announces truthfully, she expects to get
$$\frac{1}{2}\left(\frac{L}{2} - L\right) + \frac{1}{2}\left(-\frac{L}{2}\right) = -\frac{L}{2}\,.$$
If she lies, she expect to get $-H/2$; hence, she does better to tell the truth than to lie.

2. $v = (H, H)$. If a party announces truthfully, she expects to get $-H/2$ (see calculation above). If she lies, she must do the task, but is paid $H/2$; hence, her net is $-H/2$. She does least as well to tell the truth; that is, truth telling is a best response.
3. $v = (L, H)$ or (H, L). Consider the higher-cost party. If she tells the truth, she expects to get $-H/2$. If she lies, she expects to get
$$\frac{1}{2}\left(\frac{L}{2} - H\right) + \frac{1}{2}\left(-\frac{L}{2}\right) = -\frac{H}{2}\,;$$
so truth telling is a best response. Consider the lower-cost party. If she tells the truth, she expects to get $H/2 - L$. If she lies, she can expect
$$\frac{1}{2}\left(\frac{H}{2} - L\right) + \frac{1}{2}\left(-\frac{H}{2}\right) = -\frac{L}{2}\,.$$
Because
$$\frac{H}{2} - L > \frac{L}{2} - L = -\frac{L}{2}\,,$$
she does better to tell the truth.

Although such mechanisms have an artificial feel to them, they can often be converted into more realistic looking contracts. For instance, the above mechanism can be converted into an *option contract*: Party 1 is obligated to do the task and Party 2 is obligated to pay Party 1 $H/2$. However, Party 2 has the right (the option) to assume responsibility for the task if he wishes and, if he assumes responsibility, then Party 1 is obligated to pay him $L/2$. It is readily verified that Party 2 never gains from exercising his option if $c_2 = H$ and always gains if $c_2 = L$. Hence, Party 2 assumes responsibility only if he is low cost, which ensures that the task is always undertaken by the lower-cost party.

The above mechanisms are *balanced*, in the sense that any payment made by one of the parties is received by the other. The literature generally restricts itself to balanced mechanisms; in part, this is done because *un*balanced mechanisms seem at odds with what we observe in real life and, in part, because unbalanced mechanisms are too strong. The entire observable but unverifiable notion could be rendered completely meaningless, for instance, by using an unbalanced "shoot-them-both" mechanism: The judge asks the parties to simultaneously state what the observable variables are. If their reports agree, then the contract is enforced as required given the commonly reported realization of the variables. If their reports differ, then they are both shot (punished severely enough that neither side would willingly make a report that he or she knew differed from the other side's). Truth telling is an equilibrium of such a mechanism and, among the many equilibria, is arguably focal.[66]

[66] Although truth telling is a focal equilibrium given no pre-mechanism communication, it is not clear that this mechanism would work if pre-mechanism communication were permitted. That is, if you hear the other party say on the way up the court steps, "I will announce X," then you would also have to announce X even if that is neither the truth nor an advantageous statement.

4.3.2 The holdup problem and renegotiation

As discussed in 2.5.2 above, the possibility of holdup helps explain duress and related limitations on contractual freedom. The holdup problem has also been the subject of intense research in the area of incomplete contracts.[67] This recent work has focused on the extent to which contract design can provide an alternative mechanism for addressing policing holdup in situations where judicial intervention is unavailable or impractical. As such, this research can be seen as a particular application of the question, considered in the previous section, "does incompleteness matter?"

An example illustrates the problem. One firm, A, is to develop a product that a second firm, B, will market (A, *e.g.,* is a studio and B is a film distribution company). Because of its expertise, B can observe the quality of the product A has produced; yet quality could be sufficiently amorphous that it would be difficult, if not impossible, to describe in a contract or verify in court unambiguously. Setting incentives correctly would, thus, seem problematic. If B commits to buy the product at a fixed price, then A, acting opportunistically, has no incentive to invest enough in producing a quality product. If the contract is a revenue-sharing contract, then A and B's incentives could be too weak because of the teams problem (Holmstrom, 1982). Finally, if no contract is signed in advance, but the parties bargain after A has developed the product, then B could holdup A—that is, capture some of the value A creates by bargaining opportunistically—thereby adversely affecting A's incentives.

Demski and Sappington (1991) propose option contracts as a solution to this problem. Specifically, let $I \in \mathcal{I} \subset \mathbb{R}_+$ be A's investment (in dollars, say) and $\beta(I)$ be B's value of the product as a function of A's investment.[68] Assume for I and I' in $\mathcal{I}$ that if $I' > I$, then $\beta(I') \geq \beta(I)$. Suppose that I^* is the optimal investment for A to make. Demski and Sappington's solution is for the parties to sign an option contract whereby B has the right to acquire the product at a strike price of $\beta(I^*)$. If A underinvests, then B won't exercise the option, so A loses. It is a waste for A to overinvest. Hence, A should invest the appropriate amount exactly and B, being indifferent between exercising or letting the option lapse, can be assumed to exercise in equilibrium. Any desired sharing of expected surplus can be achieved through non-contingent transfers.

The Demski and Sappington solution is, unfortunately, vulnerable to renegotiation (Edlin and Hermalin, 2000). Recall that B is indifferent between exercising its option and letting it lapse, but exercises in equilibrium. Recognizing that, at the exercise date, its payoff is thus independent of its decision, B should pursue the following strategy: Let its option lapse, but then approach A about renegotiating a deal at a lower price.[69] Because there would otherwise be no rationale for trade, it must be that the product is worth less than $\beta(I^*)$ to A. Hence, there is a gain to the parties from negotiating a new price and, by the logic of the Coase theorem (see §2.2.2), they presumably will. Because the new price will be less than $\beta(I^*)$, this strategy of B's dominates a strategy of simply exercising the original option.[70] As a rational player, A will anticipate this; that is, that it will ultimately be held up despite the option contract. Anticipation of holdup will adversely affect A's incentives.

Some authors (*e.g.,* Bernheim and Whinston, 1998; Nöldeke and Schmidt, 1998), have sought to restore the capacity of option contracts to solve the holdup problem. Their approach has been primarily to allow the parties to structure the option contracts in such a manner that the contracts are robust to the sort of renegotiation described above.

Edlin and Hermalin (2000) are less sanguine than these other authors about the parties' ability to make their contracts robust to renegotiation. Accordingly, they ask whether option contracts could still work, even given the threat of renegotiation, if the strike price is lowered from $\beta(I^*)$ to a more appropriate level? If the derivative at I^* of the product's value should it remain in A's hands is greater than $d\beta(I^*)/dI$, then it is possible to achieve the efficient outcome using an option contract (with a strike price below $\beta(I^*)$). If, however, $d\beta(I^*)/dI$ is greater, then Edlin and Hermalin prove that no contract (option or otherwise) can achieve the efficient outcome. To summarize: Recognizing the possibility of renegotiation, if any contract can achieve an efficient outcome, then an option contract can; but there are circumstances in which no contract achieves an efficient outcome.[71]

The literature on holdup problems yields, therefore, a mixed message with regard to the importance of the distinction between observable and verifiable. Under a variety of conditions, the distinction is ultimately meaningless, but there are circumstances in which the distinction is important. It is worth noting that, in terms of information, this literature has considered a rather stark world—if a variable is unverifiable, then it is (often implicitly) assumed that there is no correlated signal that is verifiable. As discussed in the previous

[67] A partial list of articles in this area is Bernheim and Whinston (1998), Che and Hausch (1999), Chung (1991), Demski and Sappington (1991), Edlin and Hermalin (2000), Edlin and Reichelstein (1996), Grossman and Hart (1986), Hermalin and Katz (1993), MacLeod and Malcolmson (1993), Nöldeke and Schmidt (1995, 1998), and Rogerson (1992).

[68] $\beta(I)$ could be an expected value and it subsumes B having made the optimal marketing decisions conditional on I.

[69] In a technical sense, this is *not* renegotiation insofar as the original contract has been honored—B had the right *not* to exercise the option. B is, in a technical sense, engaging in new negotiations following the expiration of the original contract.

[70] This argument doesn't hold if the bargaining game in renegotiation gives all the bargaining power to A; see Bernheim and Whinston (1998) and Edlin and Hermalin (2000) for differing views on the feasibility of (essentially) assigning the bargaining power to A *ex ante*. In addition, it could be welfare improving for the law to adopt rules that prevent renegotiation (Shavell, 2005); although, as noted in footnote 69 *supra*, technically this bargaining between A and B is not renegotiation, but new negotiations following the expiration of the original contract.

[71] When first-best efficiency is not attainable, Edlin and Hermalin (2000) show that a simple sales contract is the second-best efficient contract. Because a simple sales contract can be replicated by an option contract with a very high strike price, there is a sense in which the optimal contract is always an option contract.

subsection, were such signals to exist, then the circumstances in which this distinction mattered would be even further circumscribed.

4.4 Legal doctrines addressing contractual incompleteness

Given the complexity of the foregoing theoretical discussion, it should not be surprising that a significant portion of the law of contracts deals in one respect or another with problems of interpretation. Such problems include: how to determine whether the parties have entered into a contract, how to determine the terms of any contract that is formed, and what to do if the contract does not explicitly or implicitly cover the situation that has arisen *ex post*. It is not feasible to review this entire body of doctrine here (for such reviews, see the sources listed in §1.3.2), but we can survey the main regulatory strategies that the law uses to deal with the economic problem of incomplete contracts. We begin by discussing three major themes that pervade the law in this area—the role of default terms, the dichotomy of form and substance, and the tension between objective and subjective modes of interpretation—and then move to a series of specific doctrinal rules that illustrate these themes in operation.

4.4.1 Contract interpretation generally

The function of default rules. In §2.1 above, we discussed the role played by contractual default rules in regulating market failures. The primary function of such rules, however, is to provide guidance for those interpreting contract terms on which the parties do not otherwise clearly agree. For example, in an ordinary sale of goods, the default rule is that the buyer must pay in cash at the time of delivery; if the parties wish to provide for credit or some other medium of payment, they must specify so in their agreement. Such default rules extend to virtually all aspects of contractual agreements, including remedial terms such as damages; and no legal system could operate without them.

The law and economics literature has taken three main approaches to the question of how default terms should be set. These approaches should be understood as reflecting different means towards achieving the goal of an efficient system for completing incomplete contracts, each of which is premised on a different assumption regarding what types of transaction costs are most empirically common.

One approach, following the work of Goetz and Scott (1985), argues that default rules should be chosen to provide terms that would minimize the cumulative transaction costs incurred by parties contracting around them.[72] In the special case where all parties face similar contracting costs and it is equally costly to contract around all terms, for instance, this implies terms that would be favored by a majority. If, conversely, some terms are costlier to contract out

[72]This argument is typically presented in intuitive form, but it could easily be formalized along the lines of §2.2.2 *supra*.

of or into than others (for example, if it is harder for the parties to specify multi-factor standards than simple rules), then other things being equal the default should be set to the terms that are easiest for the parties to escape (Schwartz and Scott, 2004). One might call this the transaction-cost-reducing approach to default rules, with the caveat that the underlying concept of transaction costs may be ambiguously defined.

A second approach recognizes that contracting costs will lead many parties to stick with a default rule even if it is not the one they would have chosen. Hence, this approach advocates choosing default terms for their substantive efficiency. Such an approach is most appealing in settings where there are substantial network externalities, adverse selection, or bounded rationality in contracting, so that individual parties are reluctant to depart from familiar and widely-used terms. For instance, Korobkin (1998a,b) presents evidence suggesting that cognitive and social-psychological factors lead parties to retain default (or standardized) contract terms that they would be better off contracting around; and Kahan and Klausner (1997) have argued that the network externalities are widespread in the corporate and business area due to the regular use of standardized forms, the value of which depends on a population of users who have invested in expertise in its application. This second approach might be called the central-command approach to default rules.

A third approach, following Ayres and Gertner (1989, 1992), argues that default rules should be used to encourage informationally advantaged parties to reveal their types before contracting, even if this means choosing terms that in equilibrium no one wants to use. They offer as an example the doctrine of *contra proferentem*, under which ambiguous contract terms are interpreted against the interest of the drafting party. Such an interpretation is unlikely to be what the drafter intended, they argue, but it serves as an incentive for the drafter to choose clear language that will convey to the counterparty the meaning of relevant terms. Ayres and Gertner, and much of the literature that follows, refer to such rules as penalty defaults, because they choose terms that operate as penalty for nondisclosure; an alternative and perhaps more precise terminology is to call this the information-forcing approach.

The concept of penalty default has been influential in the literature for its lucid illustration of the value of screening and signaling models in the analysis of contract law. Its general applicability in practice, however, is unclear from both positive and normative perspectives. From a normative standpoint, information forcing is only desirable under some circumstances; as we discuss in §3.3.3 and §5.2.7, it may be necessary to allow parties to keep the fruits of their informational advantage in order to induce them to acquire information initially. And even when it would be desirable to promote information revelation, a penalty default rule will fail to provide sufficient incentives in this regard if the value of the rents associated with informational advantage exceed the penalty. This can be the case if the informed party lacks bargaining power in a situation of bilateral monopoly (Johnston, 1990), if the information would reveal trade secrets valuable beyond the instant contract (Ben-Shahar and Bernstein, 2000), or if the asymmetric information is multidimensional and revelation along one

dimension would allow the recipient to draw inferences about another (Adler, 1999). From a positive standpoint, conversely, E. Posner (2005) has recently argued that most actual legal rules do not fit the penalty default model, basing his argument both on a survey of Ayres and Gertner's original examples, and on case law generally.

Form versus substance in contract interpretation. The key policy question underlying contract interpretation is how thorough the interpretive process should be; and this question is commonly articulated in terms of the dichotomy of form and substance. Specifically, many legal rules law have the effect of privileging certain interpretive materials and discounting others. Such rules are often termed formal or formalistic in that they confine attention to a subset of materials that may or may not give rise to the same inferences as would the universe of materials as a whole. A more substantive approach to contract interpretation, in contrast, would attempt to come to a more all-things-considered understanding, based on all of the materials reasonably available. For example, under the common law, the parol evidence rule (though subject to many exceptions) provides that a written document that integrates the parties' agreement may not be contradicted or varied by evidence of prior or contemporaneous oral understandings (E. Posner, 1998). In the absence of a writing, however, a party may introduce and a court can consider any information it wishes in order to determine the content of a contractual agreement.

From the economic viewpoint (Katz, 2004), the question of form versus substance can be assimilated to the problem of optimal information acquisition. A fuller or broader context can always be purchased, but at a cost of time, trouble, and interference with incentives, so it pays to stop at some optimal point. The problem can be formalized as follows: let the decision rule $D(\cdot)$ denote a function that maps from a set of facts to a legal outcome, for example, an award of damages. A court's *ex post* interpretation of the facts will depend upon the information available to it, which we can denote by information set $\mathcal{I}$. (For example, a court that is aware of commercial trade usage will interpret the words of a contract differently from one that is not.) This information set depends on three potential sources: the information embedded by the parties in the contract itself (denoted as $\mathcal{I}_0$), the information presented by the parties *ex post* at trial (denoted as $\mathcal{I}_1$), and the information available to the court based on its general knowledge and experience ($\mathcal{I}_a$). The information introduced by the parties in the contract or at trial is a variable that is either chosen in a cooperative game to maximize their returns from contracting or litigation, or the equilibrium outcome of a non-cooperative game in which they individually decide what information to introduce. Introducing information is costly, but the costs of *ex post* and *ex ante* information can interact, and more information can help the court reach a more efficient decision.

Within this framework, a legal interpretation regime consists of a function $R(\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_a)$ that specifies how the interpreting agent combines the three possible sources of information (for example, the most anti-formalist would be that

the agent can consider all information available, in which case R would be the union function). A more formal regime would restrict R by limiting the influence of certain categories of information; for example, under the parol evidence rule, all information in $\mathcal{I}_1$ that was inconsistent with $\mathcal{I}_0$ would be thrown out. Conversely, a regime that disfavored standardized contracts on the theory that no one reads them would throw out any information in $\mathcal{I}_0$ that was inconsistent with $\mathcal{I}_1$. The framework is quite general and can incorporate many types of evidentiary systems; for instance, in a pure adversarial system, taking to an extreme the practices of common-law regimes, the agent could not take account of any information in $\mathcal{I}_a$ that was not confirmed by information introduced by the parties in $\mathcal{I}_1$. (Of course, the rule will affect the content of the information presented; if extrinsic information contradicting the written contract is ignored, no one will want to incur resources to present it.)

Given this framework, the expected legal outcome aggregating over all interpreting agents will be $V = \mathbb{E}_a\{D(R(\mathcal{I}_0, \mathcal{I}_1, \mathcal{I}_a))\}$. The parties will choose what information to embed in the contract, and to introduce at the time of a dispute, in order to maximize their individual returns in this outcome, and the lawmaker (who could be the parties themselves specifying in the contract what interpretation rule they wish to be followed) will choose the function R in order to maximize expected contractual surplus. From the standpoint of the lawmaker, this is just a constrained principal-agent problem, the solution of which depends on the particular costs associated with $\mathcal{I}_0$ and $\mathcal{I}_1$, and the incentive properties of the decision rule D with regard to the primary behavior governed by the contract (here suppressed in the notation, and assumed to occur at an intermediate stage between the writing of a contract and the potential emergence of a dispute).

The considerations that determine the optimal approach to contract interpretation are thus quite broad-ranging. The regime of contract interpretation will influence contracting parties' behavior in many respects: with regard to decisions to breach, to take advance precautions, to mitigate damages, to gather and communicate information, to allocate risk, to make reliance investments, to behave opportunistically, and to spend resources in litigation, and so on. Accordingly, it is difficult to draw strong general conclusions regarding how interpretation should proceed.

If one is willing to make restrictive assumptions about the costs and benefits of information acquisition, however, it is possible to reach more specific conclusions. For instance, Schwartz and Scott (2004) develop a model in which parties benefit when their contract terms are interpreted correctly on average, but are relatively indifferent (due to risk-neutrality) to interpretive variance around the mean. In this case it would be optimal for courts to make interpretations on a minimum evidentiary base, because additional interpretative efforts are costly but provide no incentive benefits from an *ex ante* perspective. Conversely, Shavell (in press) assumes that *ex ante* contract-writing costs are positive and that it is possible to write at least some interpretative rules so that they can be applied costlessly (or more realistically, that the marginal cost of applying the rule is zero given that the parties are going to court anyway). In this case, it

will be optimal for the parties to leave at least some open contract terms to be filled in *ex post* by courts, and if contract-writing costs are sufficiently high, for courts actually to override the written terms in favor of another interpretation.

Obviously, such assumptions are special; in their absence, perhaps the best that can be said is that private parties should be allowed the leeway to choose their favored interpretative regime—a leeway not always recognized by the legal system. For public lawmakers, who likely lack detailed contextual knowledge about the costs and benefits of information, however, it is possible to offer some general rules of thumb. For example, formal interpretation is more efficient, other things being equal, when (1) *ex ante* negotiation costs are low relative to renegotiation costs; (2) either the parties or the tribunal are likely to be biased in interpreting the contract; (3) the chance of a dispute (or the *ex post* stakes in the event of a dispute) are relatively high; (4) legal outcomes are relatively sensitive to litigation expenditure; (5) specific investments must be undertaken by persons distant from the transaction, or have value that is relatively context-independent; (6) the parties have relatively good control over their contract-drafting agents relative to their negotiating agents; and (7) the parties have relatively good access to non-legal sanctions. In all these situations, the expected costs of filling contractual gaps *ex post* are low relative to the costs of filling them in *ex ante* so that it pays to undertake more effort in doing so gaps up front. Conversely, when renegotiation costs are high, the extent of bias is significant, and so on, the expected costs of *ex post* gap-filling are relatively low and it pays to defer more of this effort until an interpretative dispute arises. The logic here is analogous to Kaplow's analyses of the choice between rules and standards (1992) and the optimal complexity of rules (1995).

It follows from these heuristic principles that substantive interpretation is relatively more valuable to small and infrequent traders, who are less well-placed to undertake the fixed cost of detailed *ex ante* negotiation; who have relatively poor access to reputational networks *ex post*; who are likely to do their own contract negotiating but to contract out when acquiring legal services; who are less likely to be able to recover specific investments in substitute exchanges; and who possibly are less likely to face bias in *ex post* judicial tribunals. Conversely, large and experienced traders should prefer their contracts to be governed by relatively formalistic rules of interpretation. This distinction is consistent with casual empiricism—in general, it is experienced traders whom we observe contracting into relatively formal enforcement regimes through devices such as arbitration and forum selection clauses—and is also generally accepted in the case law and commentary.

Objective versus subjective interpretation. The default rule concept and the form-substance dichotomy both reflect the underlying problem of information acquisition and transmission, but focus on different dimensions of this problem. Specifically, the doctrines relating to form and substance are primarily concerned with the relationship between the contracting parties and the *ex post* interpreting tribunal; they govern the extent to which the burden of information

acquisition is allocated between the parties and the tribunal, and they affect the parties' incentives to transmit information about contractual meaning to the tribunal. Doctrines relating to default rules also implicate the relationship between parties and tribunal (especially when viewed from the transaction-cost-reducing approach of Goetz and Scott, 1985), but they additionally concern the contracting parties' informational relationship with each other (as emphasized by the penalty-default approach of Ayres and Gertner). These two dimensions of informational incentives interact in a third major theme of the law of interpretative doctrine: the tension between objective and subjective perspectives.

In this context, objective interpretation refers to the meaning that would be recognized by a reasonable outside observer, while subjective interpretation refers to the meaning as actually understood by the parties. For example, the case of *Lucy v. Zehmer* involved a dispute over whether the parties had agreed to the sale of the defendant's farm when, after a night of drinking, they signed a napkin that contained words of sale, referred to the farm, and stated a price. The defendant claimed that he was joking when he signed the napkin, and that the plaintiff either knew or should have known this. The plaintiff for his part insisted that the deal was a serious one. In the end, the court sided with the plaintiff, but rather than emphasizing the plaintiff's state of mind as its justification, it referred to publicly observable factors such as the duration of the negotiations, the existence of previous negotiations between the parties, and the reasonableness of the contract price.

In general, however, official doctrine does not give clear indication of when objective meaning controls and when subjective meaning controls. The Second Restatement of Contracts (1981), generally recognized as the most influential modern summary of American doctrine in the area, states that subjective meaning controls if the parties attach the same meaning to an agreement, or if the parties attach different meanings to the agreement and one (and only one) of them is or should be aware of the other's understanding. (For instance, if the drafter of a form knows that the form contains an unusual term and also knows that the counterparty is unaware of the presence of the term, and the counterparty is not otherwise charged with knowledge of the term, then the term is not part of the contract.) On the other hand, more traditional authorities reject these principles and insist that unreasonable meanings will not be enforced; and the Restatement itself elsewhere provides that when the parties attach different meanings to the agreement and there is no asymmetric information between them, various objective factors will be used to determine the contract's meaning.

The practical complexity of the doctrine arises from the fact that it is being used to regulate information transmission both between the parties (*i.e.,* the problem of observability), and between the parties and the court (*i.e.,* the problem of verifiability). For example, the Restatement's rule that contracts are to be interpreted according to the meaning of the asymmetrically uninformed party plainly operates as a penalty default rule, and can be justified either as forcing disclosure or as protecting the distributional interests of the uninformed. And conversely, the rule that charges parties with knowledge of meanings of

which they should be aware operates as an incentive for investigation.[73] But the objective-subjective distinction can also be understood in terms of the dichotomy of form and substance, in that an objective standard of interpretation directs the interpreter to limit its attention to factors that would be accessible to an objective observer, while a subjective interpretive standard directs the interpreter also to consider factors that might be accessible only to the parties to the contract. Thus, all the factors that bear on the optimal choice between formal and substantive interpretation (for example, the difficulty of litigating the parties' states of mind *ex post* as opposed to a counterparty's contrary meaning *ex ante* or the likely variance of error when making such assessments also bear on the choice between objective and subjective interpretation.

4.4.2 Specific interpretative doctrines

Parol evidence. In many cases, especially when substantial value is at stake, parties will memorialize their agreement in the form of a final written document. When they do, the intended effect with regard to pre-contractual communications can be ambiguous. For instance, suppose that, in oral negotiations, the buyer requests a clarification with respect to some technical feature of the goods and the seller provides clarification, but the final written document makes no mention of the issue. Should we regard the clarified term as implicitly incorporated into the final writing, or should we instead regard it as a rejected offer that the parties meant to exclude?

In common law systems, this question is addressed by a network of doctrines that are collectively referred to as the parol evidence rule. Under this rule, if the parties adopt a particular document as an authoritative statement of their contract—or in doctrinal language, as an integrated agreement—then the presumption is that they meant to reject all prior inconsistent communications. As a procedural matter, furthermore, the rule operates to exclude all evidence of antecedent understandings or negotiations that would vary or contradict the words of the document.

The parol evidence rule is subject to a number of exceptions and counter-doctrines (so much so that some commentators have suggested that it is misleading to refer to it as a rule); and the overall trend during the twentieth century has been toward decreased stringency in its application (typically, by finding, based on extrinsic evidence, that an apparently complete document was, despite superficial appearances, not intended by the parties to be authoritative). More recently, however, a number of commentators writing from an economically influenced perspective (*e.g.,* Scott, 2000; Schwartz and Scott, 2004) have argued for its re-invigoration.

Formal economic analysis of the rule, however, has been limited. Eric Posner

[73] Although, from a neoclassical economic perspective, it is unclear why it should be necessary to *provide* (strengthen) such an incentive. Presumably, private incentives should be sufficient (or even socially excessive) in this regard.

(1998) models the rule as posing a trade-off between Type 1 error (enforcing a term that is not part of the contract) and Type 2 error (failing to enforce a term that is part of the contract), as well as between transaction costs *ex ante* (*e.g.,* contract-writing costs) and *ex post* (*e.g.,* litigation costs). Posner's model should be understood as a special and more doctrinally detailed case of the general analysis of form and substance discussed *supra*. On his analysis, the rule makes little sense when courts process *ex post* information accurately and cheaply; it is justified when courts have poor ability to process *ex post* information and contract-writing costs are low. The implications for the more problematic situation where both contracting and litigation costs are high, however, are unclear and must presumably await future empirical investigation.

It is worth noting that the parol evidence rule does not apply to post-contractual communications, even though their use raises many of the same costs and benefits as pre-contractual communications do. Instead, such communications are analyzed using more flexible interpretative standards, and can operate as modifications or as waivers of contractual duty even when made informally.[74] It is possible that timing issues relating to renegotiation, holdup, and the increased value of *ex post* could justify this disparate treatment, but at present this doctrinal distinction remains undertheorized from both legal and economic viewpoints.

Trade usage and course of dealing. In contrast to the disfavored category of parol evidence, other categories of material are given specially privileged status when interpreting individual contracts. The most important of these is trade usage, which refers to any practice of dealing that is sufficiently regular and widespread in the relevant area or line of trade as to justify an expectation that it will be observed in the particular case.[75] For example, a contract for the sale of two-by-four wooden planks will ordinarily be read to refer not to planks that are literally two inches by four inches, but to the smaller $1\frac{3}{4} \times 3\frac{1}{2}$ planks that generally pass under that description in the lumber business; a buyer who wants planks that are literally two by four must state so explicitly.

This doctrine can be justified in terms of both transaction cost reduction and information forcing. From the viewpoint of transaction cost reduction, if most contracting parties wish to follow a trade usage, it saves transactional resources for interpreters to read it into all contracts, so that only those who wish to disclaim it need undertake the extra costs of negotiating explicitly. From the viewpoint of information forcing, to the extent that one party subjectively intends to contract with regard to trade usage and the other does not, these differing understandings may lead them to enter into an inefficient exchange. Given that the trade usage is well evidenced and other usages are less so, re-

[74] In some settings (*e.g.,* in sales contracts governed by Article 2 of the Uniform Commercial Code) it is possible for the parties to exclude oral modifications by using an appropriate clause in their original written contract, but it is not possible for them to exclude all possibility of *ex post* waiver.

[75] Restatement (Second) of Contract §222, Uniform Commercial Code §1-205.

quiring everyone to learn it offers a cheaper way of putting the parties on a similar informational footing.[76]

Such justifications are only persuasive, however, if trade usages can be identified cheaply and reliably, and if parties who wish to opt out of standard usage can do so at low cost. A number of recent commentators, however, have called both these assumptions into question. With regard to the first of these assumptions, Bernstein (1996, 1999) suggests that in many instances trade usages—in the sense of unwritten terms that are understood by the parties to impose actual legal obligations—do not actually exist, and that in any event, courts cannot tell the difference between behavior that evidences the existence of unwritten legal entitlements and behavior that instead represents settlement in the shadow of such entitlements. Also, Craswell (2000) has argued that courts often identify (or purport to identify) the actual content of any contested custom by consulting their own judgment, or the judgment of outside witnesses, about what custom would be most efficient. With regard to the second assumption, various commentators (*e.g.*, Scott, 2002) have argued that although contracting around trade usage is permissible under official doctrine, courts in practice imply so heavy a presumption against it so as to make trade usage (or more properly, what courts consider as trade usage) effectively mandatory.

The related doctrine of course of dealing, which refers to a sequence of conduct between the parties to an individual contract, raises similar problems. Whether a pattern of bilateral behavior amounts to a legal understanding, however, is even more difficult to determine than with group behavior because such patterns tend to be less well evidenced and because the participants may wish to recharacterize their past behavior with an eye toward influencing the outcome of a current dispute. In addition, as Ben-Shahar (1999) has argued, allowing course of dealing to influence the interpretation of future contracts can lead parties to be excessively rigid when demanding contractual performance, for fear that greater flexibility will prejudice their future rights.

Such functional problems can be understood as stemming from the basic problem of contractual incompleteness, with the extra twist that it is not just the individual contract terms that are unverifiable, but also the alleged trade usage or course of dealing. Accordingly, whether courts should read trade usage and course of dealing into contracts as a default rule, or whether parties should make greater efforts to disclaim such a reading, remain open questions.

Change of circumstance. The doctrines of trade usage and course of dealing are used to fill out incomplete contracts in routine circumstances. In contrast, another set of doctrines are used to imply missing terms in unusual circumstances—specifically, when unforeseen events intervene to reduce substantially one or both parties' gains from trade under the original contract. For example, in *Taylor v. Caldwell*, a landlord whose building was destroyed in a fire was excused from liability to a promoter who had leased the premises for

[76] See Hermalin (2001) for a brief survey of research on the efficiency gains to be had from terminological conventions.

a theatrical event, even though the lease contract contained no explicit excuse of this sort; and in *Krell v. Henry*, a royalist socialite who had leased, at inflated short-term rates, a flat overlooking a public square in which the incoming British king was to be crowned was excused from the obligation to pay rent after the coronation was postponed due to the king's emergency appendectomy.

When this principle is used to excuse a producer or seller, it is often referred to as the doctrine of impossibility, although this term is a misnomer in that (1) the doctrine is often applied to situations in which it is highly burdensome but not impossible for the seller to perform, and (2) even if performance is not physically possible, the payment of damages is. Other labels include impracticability, commercial impracticability, and (especially when the buyer is excused, as in *Krell v. Henry*) frustration of purpose.

An analogous principle is sometimes used to excuse parties from complying with conditions that turn out to be unexpectedly burdensome, so long as the excuse does not unduly disadvantage the counterparty. (For example, a condition in an insurance contract requiring all claims to be made within two weeks of the occurrence of a casualty would typically be excused if the delay was caused by the insured party's being incapacitated in the underlying accident.) Finally, as indicated in §1.3.2 above, parties are sometimes excused from liability under the doctrine of mistake when they contract without the benefit of information that, once revealed, substantially alters the contract's efficiency or distribution of surplus. As observed above, from an information-theoretic perspective this situation is functionally equivalent to the case of *ex post* changed circumstances.

Economically influenced commentators have most commonly justified such doctrines on transaction-cost-reduction grounds. Specifically, Posner and Rosenfield (1977) suggest that, insofar as the doctrines apply only to excuse a party who neither has hidden information nor takes a hidden action with regard to the excusing event, they mimic the risk allocation terms that would most likely be provided in a complete contingent contract. (For instance, in *Taylor v. Caldwell*, given that the theater owner had already suffered the loss of a building, he was unlikely to be the least-cost insurer of the promoter's losses, at least in the absence of any specific facts evidencing moral hazard or adverse selection.)

The changed-circumstances doctrines can also be understood as special applications of the penalty-default and form-versus-substance frameworks outlined above. From a penalty default perspective, a contingency that could reasonably have been anticipated *ex ante* by one of the parties should not count as an excuse, because denying the excuse will encourage the informed party to bring the issue into the contractual negotiation. Conversely, from a form-substance perspective, contingencies that are easier to specify and ascertain *ex post* either because information is better after the fact or because their probability is too low to warrant up-front attention, are appropriately left off to be dealt with by a court in the unlikely event that they arise.

It should be noted that the changed-circumstance doctrines can also be understood as special applications of the theory of optimal contract remedies along the lines of the discussion in §5 below. For instance, White (1988) points out that excusing a party for changed circumstances has consequences not just

for risk allocation, but also for various countervailing considerations such as incentives for precaution and relational investment. She argues that it is only in rare cases that zero damages are second best, and accordingly that changed-circumstance cases should be treated like any other contractual breach. On the other hand, once the costs of legal enforcement are taken into account, excuse from liability may be warranted as a litigation-saving device to be applied in extreme situations where it is sufficiently clear that the contract should not be performed and that breach is not due to any failure of precaution or disclosure (in the same way that negligence liability in tort saves administrative costs by requiring litigation only in case of a breach of duty).

Standardized contracts. Most contracting parties do not negotiate individual contracts for each occasion on which they enter into exchange; instead they use standardized forms that incorporate customary or boilerplate terms, and negotiate only over those few terms (such as price and quantity) that are essential for their transactional purposes. From an economic viewpoint, such behavior is a straightforward response to scale economies in the production of contractual terms. In legal circles, however, standardized contracts are regarded with more ambivalence. While most legal commentators accept that such contracts are a practical necessity, many continue to regard them as problematic on grounds of autonomy (because most people do not read or understand other parties' standard forms and hence cannot knowingly assent to their terms) or distributive justice (because the party who drafts the form is deemed to hold an inequitable degree of bargaining power).

At an abstract level, standardized contracts are interpreted using the same methods as contracts more generally. For instance, if the drafting party knows that a particular term in the contract is unusual, that the non-drafting party is unaware of the term, and that the non-drafting party would not enter into the contract if aware of the term, then the term is not part of the contract; as observed in the discussion of objective versus subjective interpretation above, this result follows from ordinary principles regarding subjective interpretation. At a practical level, however, most lawyers understand form contracts to constitute a special category subject to specialized rules of interpretation, such as the doctrine that contracts are to be construed against the interests of the drafter. Conventional legal wisdom holds that form-contract terms are less likely to be enforced than individually negotiated terms, especially when they are unusual or thought to be especially burdensome to the non-drafting party.

There is relatively less reason to be concerned about standard terms on market power grounds, notwithstanding many legal commentators focus on this issue. Because standardization lowers the per-unit cost of contracting, both competitive firms and those with market power will find it of value. While firms with market power might distort non-price terms from the efficient level, the direction of distortion is *a priori* ambiguous (Spence, 1975). If all consumers have the same willingness to pay for contractual terms, a monopolist will do best to provide optimal terms and to extract available profits through a high

price; and if consumers differ in their willingness to pay for contractual terms, super-optimal terms are as likely as suboptimal ones. While oligopolists could use standard forms to collude on non-price terms, it is similarly unclear in which direction their incentives cut. Accordingly, to the extent standard form contracts raise competition concerns, these are best policed by antitrust law, rather than by common-law courts applying the law of contract.

Standard form contracts do raise informational concerns, however, since they can vary substantially in their terms and the drafting party knows much more about the terms than does the non-drafting party. If contract reading is costly, parties may rationally assume that a particular contract contains the average terms available on the market, with the result being adverse selection along the lines of Akerlof's market for lemons (see also the discussion in §2.3.4). In addition, as Katz (1990b) has shown, if contract reading is a relation-specific investment (as it will be if different drafters offer different terms), drafters will be tempted to expropriate this investment by choosing terms that are just favorable enough for the non-drafter to accept after having read them, thus undermining any incentive to read. Interpretative rules that make it harder for drafters to enforce unusual standard terms without calling them to the other party's attention, such as the doctrine of *contra proferentem* or the rules requiring warranty disclaimers to be conspicuous, help to address this informational asymmetry, by putting the burden of communication on the party who can undertake it most cheaply. Such rules provide perhaps the clearest illustration of the penalty default approach discussed above.

Implied duty of good faith. Finally, most modern-day courts will require the parties to a contract to exercise good faith in the performance and enforcement of their contractual duties. Commentators disagree on exactly what this obligation requires, but as a general matter it prohibits opportunistic actions that, while complying with the bare letter of the contract, depart from it in spirit by operating to deprive a counterparty of the reasonably expected benefit of its bargain. For example, it has been held a breach of good faith for a buyer of goods to seize on an otherwise trivial defect in delivery in order to escape an unfavorable contract in a falling market, for an employer to exercise its termination rights on an at-will labor contract just before the employee's accumulated sales commissions were due to be paid, or for a realty purchaser to prevent his seller-broker from acquiring the property to be delivered under the contract (and thus forcing the seller into breach) by showing up at an initial land auction and outbidding him.

In theory, the parties could guard against opportunistic exercises of discretion by providing specific limits in their original contract. For all the reasons discussed above in §4.2, however, they often do not do so, leaving courts with a disagreeable choice between imposing additional *ex post* limits on an *ad hoc* basis, and countenancing an unanticipated and perhaps unfair result. Traditional common-law courts often responded to this dilemma by refusing to enforce such agreements entirely, on the theory that such agreements were so one-sided as to

lack formal consideration. Modern courts, in contrast, are more willing to treat such contracts as genuine bargains, and to police party opportunism directly and substantively after the fact. As our earlier discussion of form versus substance indicated, such an approach could serve the parties' economic interest if the courts' costs and error rate in filling contractual gaps *ex post* are sufficiently low, or if the parties' costs of dealing with them *ex ante* are sufficiently high.

Whether the duty of good faith is best viewed as a rule of interpretation, or alternatively as a limitation on contractual freedom, has not always been clear from the case law or commentary. The fact that the scope of the duty depends on other terms of the agreement, as well as on the overall commercial context, suggest that it is an interpretative rule; but the fact that the parties are not permitted to disclaim it (though they are permitted to stipulate its content so long as they do not do so in a "manifestly unreasonable" way) suggests that it is a regulatory intervention. The uncertainty is reinforced by the murky boundaries of the duty, which in limiting cases appears to shade into other mandatory doctrines such as unconscionability and public policy.

What counts as opportunistic behavior, however, typically lies in the eye of the beholder, and critics of the modern duty of good faith (*e.g.,* Schwartz, 1992; Bernstein, 1999; Scott, 2000) argue that the behavior that it seeks to induce is typically unverifiable, and thus not effectively subject to judicial oversight. Courts' attempts to enforce good faith, accordingly, are as likely to depart from contractual expectations as to enforce them, in the end just resulting in an increase in the cost of litigation and in the variance of judicial outcomes. On this view, enforcement of such "soft" contractual expectations should be left to private enforcement mechanisms such as reputation and repeat dealing. For the most part, however, the courts have not yet accepted this critique.

4.5 Overall assessment of the law of contract interpretation

In general, the economics of contract interpretation is a relatively unmined field compared to the economic analysis of contract law generally. While the general contours of legal understanding in this area are more or less consonant with the main insights of the economic theory of contract, the rapid and recent development of the economic literature has not yet been matched by a corresponding growth in legal scholarship. It is likely, accordingly, that this set of topics will draw increasing attention from law and economics scholars in upcoming years.

5 Enforcement of contracts

As we observed at the outset of this chapter, the rationale for contracting is to lock in a commitment *ex ante* that one or both parties would otherwise not wish to honor *ex post*. The use of a contract to establish such commitment is undermined, of course, if the contract will not be enforced in the way the parties anticipate.

The enforcement of contracts is often the province of courts, who impose sanctions on parties who violate or breach a contractual obligation. As we discuss below, contracts can also be enforced by the parties themselves, through the use of various self-help remedies. We begin, however, with a discussion of judicial enforcement, partly because this has been the focus of most of the existing literature, and partly because an understanding of judicial enforcement is useful in order to understand when and why private enforcement might be more attractive.

5.1 General issues in enforcement

5.1.1 Remedies and contractual incompleteness

In one sense, every legal dispute is about remedies. Even when a court rules that no contract was ever formed, that ruling can be seen as saying that the complaining party is not entitled to any remedy for breach of contract. Moreover, such a ruling might still entitle the parties to some other legal remedy—for example, in some circumstances, one or both might be required to return any advance payments that had been made. From an economic standpoint, what ultimately matters is not the doctrinal question, "was a valid contract formed?" but rather the remedial question, "how much money can I now collect?" (or "how much money will I now be required to pay?")

In another sense, though, the identification of certain actions as a breach of contract, and the identification of certain payments as remedies, is an artifact of contractual incompleteness. After all, in a hypothetical complete contract, the contract itself would specify the payments and other actions that would be required in every possible state of the world. With such a contract in place, there would be no reason to label some actions in some states of the world as complying with the contract, while labeling other actions in other states as noncomplying or breaching. There would also be no reason to label some (but not all) of the associated payoffs as "remedies for breach."

As discussed in §4.2 above, however, complete contingent contracts are rarely observed in the real world. In particular, real contracts are often incomplete in two ways that are relevant here. First, rather than specifying every possible action that a party might take, they may specify only those actions by each party that will suffice to comply with the contract—for example, "seller to deliver fifty widgets by July 1; buyer to pay $1,000 by July 10." Second, they may fail to specify the steps that should be taken if either party fails to take a complying action—for example, if the seller delivers only forty widgets, or delivers them on July 15; or if the buyer refuses to pay.

Parties sometimes do provide explicit terms that specify the remedy for some forms of noncompliance; these terms are known as "liquidated damage clauses" and they will be discussed in §5.3.4. In other cases, however, where the parties' contract fails to specify the consequences of noncompliance, the legal system will supply a *default* remedy, just as it supplies default rules for other questions that a contract might not address (see §4.4). As most of the literature has focused

on these default legal remedies, that is where we begin.

5.1.2 Overview of default remedies

In some cases, parties who fail to comply with their contracts may be ordered by courts to perform—in legal terms, specific performance (see §5.3.3 below), with the order backed up by threats of more severe sanctions for contempt of court. In most cases, though, the default remedy for breach has the breaching party pay money to the aggrieved party.

Following Fuller and Perdue (1936, 1937), it has become customary to identify the most common remedies as expectation damages, reliance damages, or restitution damages. *Expectation damages* attempt to put the injured party in as good a position as he would have been in if the contract had been performed. For example, if a buyer contracts to pay a price p for a good that has a gross value of v to him, but whose value can be realized only if the buyer spends r to make use of the good, full performance of the contract would leave the buyer with a net value of $v-r-p$. If the seller then breaches the contract, by failing to deliver the good, expectation damages would require the seller to pay enough to leave the buyer in that same net position: $v-r-p$. For example, if the buyer had already spent both p and r, expectation damages would require a payment of v.

By contrast, *reliance damages* require the seller to pay only enough to leave the buyer with the level of utility he would have enjoyed if the contract had not been signed. For example, if the buyer in this case would have had a net gain of zero had he not signed this contract, reliance damages would require the seller to pay enough to bring the buyer back to that position. So, for example, if the buyer had already spent both p and r, reliance damages would then equal $p+r$, enough to bring the buyer back to zero.

Another standard remedy, *restitution*, allows both parties to recover the reasonable value of whatever they gave to the other party. In this example, restitution would allow the buyer to recover damages of p.

As this example shows, these three remedies can often be ranked by size, with expectation damages usually providing the largest remedy and restitution giving the smallest. (We may assume $v > p + r$ because otherwise the buyer would not sign the contract.) Similarly, reliance usually exceeds restitution, because the buyer's reliance typically provides little benefit to the seller. Of these three remedies, expectation damages are conventionally regarded as the predominant measure in both theory and practice, with reliance and restitution reserved for cases in which there is some defect in the bargain or in the plaintiff's ability to prove it, or when equitable considerations justify a departure from the ordinary rule. In practice, however, this generality does not always hold, due to measurement difficulties and to other aspects of the legal rules governing the availability of each remedy.

For example, expectation damages entitle a plaintiff to compensation for profits that would have been earned on the breached contract, but calculating hypothetical revenues and expenses on a cancelled project is often a speculative

endeavor. Accordingly, courts often presume that the future profits to be earned on the contract approximate the costs so far incurred, thus using reliance damages as a proxy for expectation damages. Conversely, reliance damages should in principle compensate a plaintiff for the loss of forgone alternatives, but such opportunity costs are often difficult to verify *ex post*. As a result, courts often presume that the forgone opportunity would have yielded the same profits as the breached contract—in which case expectation serves as a proxy for reliance. Similarly, it is often difficult to verify the value of restitution, especially when it takes the form of services not traded on any market. Thus, courts sometimes measure restitution in terms of the reliance costs incurred by the non-breacher, though their willingness to do so may depend on their perceptions of relative fault. There are even situations in which the non-breacher is permitted to choose among expectation and restitution, or among expectation and reliance; obviously, in those cases the more generous remedy will be chosen.

In addition, as we discuss below, each of these remedies can be modified or adjusted in various ways: for example, if the victim could have prevented some or all of the losses by appropriate mitigating behavior. As we also discuss below, there are a few additional remedies that are occasionally imposed following a contract breach, including punitive damages in rare cases.

In part for these reasons, it is often more helpful to think of a monetary damage remedy as a continuous choice variable d, whose value can in principle be set by the legal system to any level. In some cases, the damages awarded may more appropriately be written $d(\mathbf{x})$, where $\mathbf{x}$ is a vector representing any number of actions taken by one or both of the parties. Put less formally, remedies for breach can affect the parties' incentives both indirectly, just by threatening a party with having to pay some amount d in certain states of the world; and more directly, by conditioning the size of d on some particular behavior x that the law seeks to influence. In general, though, damage awards that are conditioned on particular actions will often be more difficult for courts to implement, depending on the ease with which those actions can be verified by courts.

5.2 Monetary damages for breach of contract

We now turn to the effects of monetary remedies of various sizes and descriptions. Obviously, larger monetary remedies increase the non-breacher's payoff (and reduce the breacher's payoff) in the event of a breach, while smaller remedies produce the opposite effect. However, this effect by itself is merely distributional, and will not by itself change the transaction's expected value. As long as both parties correctly estimate the probability of a breach, the prospect of liability for higher (or lower) damages can be offset by charging a higher (or lower) price, leaving both parties with the same expected return.

Instead, changes in the expected value of the transaction must come from some other cause. For example, if one or both parties are risk averse, there can be welfare gains from shifting more of the variance in payoffs to the less risk-averse party. We discuss this possibility in §5.2.9 below.

More important, in most cases the expected value of the transaction depends

in part on various actions the parties might take to increase (or reduce) that value. Moreover, the remedy for breach will usually affect the parties' incentive to take these actions—for example, their incentive to perform the contract rather than breaching, or to take precautions against contingencies that might lead them to breach, or to make transaction-specific investments whose value will be lost if there is a breach. Each of these incentive effects is discussed below.

5.2.1 The breacher's decision to perform or breach

Suppose that a seller must decide whether to perform a contract, at a cost c, when performance will confer on the buyer a value v. Total welfare is maximized if the seller performs when and only when $v \geq c$. Otherwise, it would be more efficient for the seller not to perform, an outcome often referred to as "efficient breach."[77]

To be sure, if v and c are known at the time of contracting, the parties would have no reason to contract unless v exceeded c. In many situations, however, c or v (or both) are stochastic variables whose values will not be realized until just before performance is to take place, long after the contract was signed. Accordingly, one problem of interest is to design remedial rules that will give the seller an incentive to perform the contract if and only if $v \geq c$.

As the earliest contributions to this literature recognized, a remedy of expectation damages creates exactly this incentive, as long as those damages are accurately measured.[78] As defined above, expectation damages force the seller to internalize whatever losses the buyer suffers from the seller's breach. Consequently, the seller has an incentive to breach only if her gains from breach exceed both parties' losses. (A symmetric analysis, which we omit here for the sake of brevity, could be applied to the buyer's incentives to perform or breach.)

Before turning to other relevant incentives, four points should be made concerning even this simple result. First, while the expectation remedy forces the seller to internalize the buyer's losses from breach, it does not necessarily give the seller any incentive to consider effects that her performance or breach might have on third parties. As with most of the literature, we restrict attention to cases where third-party effects are absent (but see §2.3.1 and §5.3.4).

Second, like any other remedy, expectation damages will affect the seller's incentives only to the extent that the seller actually expects to pay those damages. If, instead, there is a significant chance that the seller could escape having to pay, the seller's incentives would be correspondingly reduced. As we discuss

[77] The term "efficient breach," though widespread in the literature, is somewhat of a misnomer, and has had the unfortunate effect of leading legal scholars untrained in economics to suppose that there is a tradeoff in this regard between efficiency and the deontological value of promise-keeping. From the perspective of a hypothetical complete contract, calling off performance when $v < c$ (possibly combined with some side payment) is precisely what the parties would have wished to specify, so a legal default rule that results in that outcome is better understood as promoting the parties' promissory intentions than subverting them. The term "efficient implied cancellation option" would be more accurate, if less arresting.

[78] See Birmingham (1970) and Barton (1972). The first formal model is Shavell (1980).

below (§5.4), the costs and other difficulties of pursuing a lawsuit may sometimes make this legal remedy less effective.

Third, if courts instead award some higher or lower measure of damages—or, equivalently, if they attempt to award expectation damages but predictably err in measuring those damages—the seller's incentives to breach will be altered. If courts tend to award damages that are higher than the true measure of expectation damages, the seller will have an even stronger incentive to avoid breaching: too strong an incentive, relative to the "efficient breach" condition described above. If courts instead tend to award lower damage measures, as is widely believed to be the case, the seller will have a weaker incentive to avoid breaching. The possibility of judicial error is particularly likely if some or all of the buyer's benefits from performance are either unobservable or nonverifiable (see §4.2 above).

Indeed, if courts instead could observe every relevant variable without error, it would be trivial to create efficient performance incentives simply by having the courts evaluate the efficiency of every breach *ex post*. If the court decided that a breach was efficient, the breacher could be excused from any remedy (or even rewarded with a bonus), while if the court decided a breach was inefficient, the breacher could be hit with huge sanctions. Viewed from this standpoint, the potential benefit of using expectation damages to create efficient incentives for performance or breach is that this remedy does not require courts to be able to evaluate the efficiency of any particular breach. The expectation remedy does, however, require courts to be able to calculate the amount the buyer would have gained from performance. How often courts are able to make such calculations is a matter of some dispute.

Fourth, and most important, the incentives to perform or breach may not even matter as long as the parties can renegotiate after the values of v and c have been realized. If this *ex post* renegotiation is possible, then—regardless of the legal remedy for breach—there should always be an agreement that will maximize both parties' gains by performing the contract if but only if performance is efficient (Shavell, 1980). To be sure, renegotiation may entail positive transaction costs, but most legal remedies (including expectation damages) also entail positive transaction costs. While some of the early literature attempted to identify the legal remedy that would satisfy the "efficient breach" condition with the lowest possible transaction costs,[79] empirical evidence on actual transaction costs is largely nonexistent, and that literature is best described as inconclusive.

As discussed earlier, the possibility of *ex post* renegotiation also plays an important role in the literature on incomplete contracts (see §4.2.6 *supra*). This similarity should not be surprising—for as noted earlier—remedies for breach are really just one aspect of contractual incompleteness. As we discuss below, the possibility of *ex post* renegotiation also complicates the economic analysis of other incentives created by legal remedies.

[79] See, *e.g.*, Schwartz (1979), Macneil (1982), and Bishop (1985).

5.2.2 The breacher's decision to take precautions

When the seller's costs c are stochastic, the distribution from which c is drawn might be given by nature, entirely independent of any action by the seller. More commonly, though, that distribution is itself a function of the seller's investment in the transaction. For example, a seller's costs might depend in part on how often her assembly line malfunctions; and the probability of such a malfunction might depend on how much the seller spent on maintenance. Typically, the seller must choose her expenditure on maintenance at one point in time; only then, after that expenditure has been made, the actual value of c will be realized.

Expenditures such as these are often referred to as "precautions" (see generally Cooter, 1985). The optimal expenditure on precautions can be defined straightforwardly in terms of (a) the cost of each possible precaution, and (b) its marginal effect on the distribution from which c is drawn, together with the welfare associated with each possible realization of c, taking account of the fact that some of those realizations will result in the contract being performed, while other realizations will result in the contract being breached. Given the usual assumption of diminishing marginal returns, there will be a unique expenditure on precautions that maximizes the total expected value of the transaction.

Interestingly, expectation damages (if measured accurately) can give sellers an incentive to choose this optimal expenditure on precautions (Kornhauser, 1983). As discussed above, expectation damages force the seller to internalize all of the buyer's losses from any realizations of c that result in a breach. As a consequence, expectation damages can in principle optimize both of the incentives discussed so far: (a) by giving the seller an incentive to choose the optimal expenditure on precautions; and (b) once the actual value of c is realized, by giving the seller an incentive to choose between performing and breaking the contract. Again, any remedies that are systematically less than expectation damages—whether by design, or because of measurement error—should reduce both of these incentives, while any remedies that systematically exceed expectation damages should increase both incentives.

To be sure, if courts could observe the seller's actual expenditures on precautions, and if they could also observe all the factors necessary to calculate the optimal expenditure, there would then be other ways to give sellers an incentive to choose the optimal level. For example, courts could excuse from liability any seller whose precautions were optimal, in much the same way that negligence standards in tort law excuse defendants whose precautions were optimal. Alternatively, if the contracting parties could specify in their contract the optimal expenditure on precautions, that requirement could itself be enforced by a court, as long as the court could verify the seller's actual expenditure. Thus, just as in the case of incentives for performance (§5.2.1), the case for using expectation damages to optimize a seller's incentives to take precautions rests in part on the assumption that it is easier for courts to observe only those factors necessary for the calculation of expectation damages (*i.e.,* only those factors that go to the value that the buyer would have received if the contract had been performed)

than it is for courts to observe any of these other factors.[80]

Alternatively, if buyers can observe a seller's precautions (before contracting), competition among sellers could give them a market incentive to choose the optimal expenditure on precautions (Kornhauser, 1983). Even if sellers do not choose their precautions until after contracting, optimal market incentives could be possible if buyers can observe sellers' reputation for taking precaution. In §5.4 below, we consider reputations and competitive markets in connection with enforcement by means other than legal remedies.

5.2.3 The non-breacher's reliance decision

The value the buyer places on performance by the seller is often a function of the buyer's investment in the transaction. For example, a business that is buying a new machine may get more value from that machine if it spends money training its employees to use it. Expenditures such as these are often referred to as reliance on the contract.

By definition, reliance expenditures are at least partially transaction-specific—for example, training designed for one particular machine may be worthless if that machine is not delivered. Although there can often be varying degrees of performance, for concreteness assume that there are only two discrete possibilities—either the contract is performed, or it is breached. Let $v(r)$ represent the value that the buyer will receive from performance conditional on his reliance investment r, while $w(r)$ represents his value if the contract is not performed. If performance of the contract is certain, the optimal reliance expenditure is simply the value that maximizes $v(r) - r$. In the more usual case, where the contract should be performed in some states of the world only, the optimal reliance expenditure is the value that maximizes

$$qv(r) + (1 - q)w(r) - r\,, \tag{9}$$

where q is the probability that contract should be performed (here taken to be independent of r; *e.g.,* with probability $1 - q$ some "act of God" makes performance prohibitively expensive for the seller).

The remedy of expectation damages ($d = v(r) - w(r) - p$, in this example) will generally not give the buyer the right incentive to choose the optimal value of r. Expectation damages are a "full insurance" remedy that gives the buyer the same net return in every state of the world; that is, the buyer's net return is $v(r) - p$ under performance and $d + w(r) = v(r) - p$ under non-performance. Unless $v'(r^*) = w'(r^*)$, where r^* maximizes expression (9), the buyer will choose a non-optimal level of reliance. If one assumes, as is often true, that the marginal return from reliance investment is greater given performance than non-performance (*i.e.,* $v'(r) > w'(r)$), then buyers have excessive incentives to rely under expectation damages.

Reliance damages ($d = r - w(r)$, in this example) also will fail to provide the buyer the proper incentives. If $v'(r) > w'(r)$, then $w'(r^*) < 1$; in this case,

[80] For a discussion of this aspect of expectation damages, see Cooter (1985).

reliance damages will encourage the buyer to rely excessively because increased expenditures on r will increase the damages they can collect.

More generally, the probability q is endogenous insofar as the level of damages (either expectation or reliance) affects the probability that the contract will in fact be performed (see §5.2.1 and §5.2.2). Even with an endogenous q, however, neither reliance nor expectation damages will generally provide the proper incentives. To see this, observe that the first-best solution requires performance whenever $v(r) - c \geq w(r)$ and, therefore, that investments maximize

$$\int_0^{v(r)-w(r)} \big(v(r) - c\big) f(c|s) dc + \int_{v(r)-w(r)}^{\infty} w(r) f(c|s) dc - r - s\,, \qquad (10)$$

where s is the seller's investment in precaution and $f(c|s)$ is the density of cost given such investment. To replicate the first-best solution with regard to whether the seller performs, it must be that $p - c \geq -d$ if and only if $v(r) - c \geq w(r)$; hence, $d = v(r) - w(r) - p$, which accords with expectation damages, but not reliance damages. However, as we saw above, expectation damages causes the buyer to face the maximization problem $v(r) - r$ with respect to his choice of reliance. As discussed, the r that solves the buyer's problem will differ from the r that maximizes social surplus (*i.e.,* expression (10)).[81]

If and only if *ex post* renegotiation is impossible, a remedy of no damages at all—*i.e.,* a regime of no liability for breach—can optimize the buyer's reliance incentives in a second-best sense (given the probability of performance), because such a regime forces the buyer to bear all of the downside as well as all of the upside of any reliance expenditure (Shavell, 1980). To be sure, such a remedy does little to optimize the seller's incentives, as discussed earlier in §5.2.1 and §5.2.2. But if the breach probability (*i.e.,* q above) is exogenous or the seller's incentives can be optimized independently (*e.g.,* through market-reputation effects), a zero-damage remedy could be optimal. This is not wholly surprising, as this forces the buyer to face the social planner's problem as given by (9) above.

Mathematically, zero is just a constant and, more generally, any constant-damage measure—that is, any measure whose value does not change with the buyer's level of reliance—will also optimize the buyer's reliance incentives (for any given probability of performance), as long as *ex post* renegotiation is still impossible.[82] Under a constant-damage measure, the buyer will still capture all of the marginal benefits, and bear all of the marginal costs, of higher or lower reliance expenditures. Moreover, a constant-damage measure could also optimize the seller's incentives, if the constant is set equal to the value that performance would have to a buyer who in fact relied optimally: $d = v(r^*) - w(r^*) - p$, in this example. In addition to optimizing the buyer's reliance expenditures, this

[81] See Shavell (1980) for greater details. Consider also Shavell (1984), Kornhauser (1983), and Rogerson (1984).

[82] Shavell (1980) considers a model in which restitution is a constant-damage measure.

will also give the seller optimal incentives, both to perform or breach and to choose an optimal level of precautions.[83]

However, as noted, this result holds only when *ex post* renegotiation is impossible. When renegotiation is possible, reliance by the buyer can have the additional effect of altering the terms likely to be reached in any renegotiation, thus further distorting the buyer's reliance incentives (recall the discussion of holdup above in §4.3.2). Under certain conditions, holdup concerns may bias the buyer's incentives toward choosing too little reliance (Rogerson, 1984).

Edlin and Reichelstein (1996) show that, in some circumstances, the parties may be able to choose contract terms that balance the risk of holdup against the risk of moral hazard, thus implementing a first-best solution (recall, too, the discussion in §4.3.2). For example, suppose the parties enter into a fixed-price contract in which they can freely adjust the quantity term, and again, suppose that the buyer must rely before he knows whether the seller will perform. If the contract quantity is set at zero (*i.e.,* if there is no enforceable contract at all), the buyer will under-rely because part of the value created by reliance will be expropriated through *ex post* holdup. On the other hand, if the contract quantity is set at the level Q^* that the buyer wishes to purchase, he will be fully insured against non-performance, and will over-rely. By continuity, there exists some contract quantity $\hat{Q} \in (0, Q^*)$ that will provide optimal reliance incentives. In the event that the seller can perform, the parties can then make up the difference between $\hat{Q}$ and Q^* *ex post*, entering into a spot transaction for an additional amount $Q^* - \hat{Q}$.

Alternatively, buyers' reliance incentives can always be optimized if courts are capable of evaluating directly the efficiency of any reliance expenditure. If courts can evaluate reliance expenditures directly, then it is easy to create optimal incentives simply by rewarding any buyer that relies optimally or by penalizing any buyer who fails to rely optimally. Indeed, there are various legal doctrines that might be interpreted as having this effect by, for example, awarding damages only for "reasonable" reliance expenditures (Goetz and Scott, 1980), or by calculating recovery based on the reliance that would have been "reasonably foreseeable" (Cooter, 1985). Obviously, the effect of any such mechanisms depends on how courts define a "reasonable" (or "reasonably foreseeable") level of reliance. This, in turn, depends heavily on the verifiability of the factors that define the optimal level of reliance (the $v(\cdot)$ and $w(\cdot)$ functions, in the model sketched here).

Of course, if the parties themselves can write a complete contract, reasonable expenditures on reliance could be spelled out in the contract itself. As noted earlier, though, this much contractual completeness is sometimes difficult or impossible. Alternatively, if parties can acquire reputations for relying optimally (or for relying excessively), then the competitive advantages of such a reputation could also induce the parties to make optimal reliance investments. We defer discussion of these possibilities to the section on non-legal enforcement (§5.4).

[83] See Cooter (1985) and Cooter and Eisenberg (1985).

5.2.4 The non-breacher's precautions

A closely-related issue concerns precautions that the buyer may be able to take to reduce the loss he will suffer if the seller breaches. For example, if there is a significant chance that the seller may not deliver the machines she promised, it might be efficient for the buyer to keep his old machines as back-ups rather than getting rid of them as soon as the contract is signed, even if keeping the old machines around is costly. If the seller does perform the contract, the expenses involved in keeping the old machines will have been unnecessary—but if the seller fails to perform, those expenses will have been well spent. Conceptually, then, *failing* to take such a precaution can be thought of as another form of reliance on the promise of performance (Cooter, 1985).

As a consequence, the legal remedy for breach can also affect buyer's incentives to take these sort of precautions. As discussed above, simple expectation or reliance remedies create a moral hazard problem that can leave buyers with incentives to rely too heavily—or, in this context, to take too few precautions. There are, however, legal doctrines that might correct that problem by limiting buyers' ability to recover for losses that could or should have been avoided by appropriate pre-breach precautions. The doctrine of *Hadley v. Baxendale*, for example, may sometimes be used to deny recovery of losses that the buyer could have prevented, by ruling that such losses were not "reasonably foreseeable" to the seller (*e.g.,* R. Posner, 2003b, p. 127). Similarly, the implied excuse doctrines—impracticability, frustration, and mistake—could in principle be used to optimize buyer's incentives by releasing the seller from liability entirely, thereby making the buyer bear all losses suffered from the seller's nonperformance in just those cases where the buyer should have taken more precautions than he did (Posner and Rosenfield, 1977). If courts release sellers from liability only when buyers have behaved suboptimally, in a manner similar to a contributory negligence rule in tort law, that could give buyers an incentive to optimize their own behavior. And as long as buyers optimize their own behavior, courts could then return to holding sellers fully liable, thereby (in principle) optimizing both parties' incentives.

However, these doctrines can have this effect only if courts are able to evaluate fully the precautions available to buyers *ex ante*, in order to recognize cases when the buyer should indeed have taken additional precautions. The extent of courts' ability to do this—and to do it reliably enough so that buyers will know when they will be held liable if they do not take precautions—is a matter of some dispute.[84]

5.2.5 The non-breacher's mitigation of losses

Another common form of precaution involves steps the buyer takes after the seller's breach becomes final. For example, if the buyer has bought special equipment to use with a machine that is never delivered, that equipment can

[84] For a skeptical view, see Kull (1991).

sometimes be salvaged or re-used, though perhaps not for its full original value. In the notation used in the preceding section, this salvage value was implicit in the $w(r)$ function, but that notation suppresses the fact that the buyer (the non-breacher) may have to take various steps in order to realize that salvage value. Moreover, some such steps are likely to be efficient while others are not—for example, in some cases the cost that must be incurred to salvage unused equipment may exceed the equipment's salvage value.

Courts are rather more willing to consider such *ex post* opportunities for mitigation of losses and to require buyers to take advantage of them than they are to consider the precautions that buyers might have taken before breach. In part this is because the value of mitigation is especially salient after the fact of breach (or from an incomplete contracts perspective, more easily verifiable). As a matter of law, this issue is most explicitly addressed by the mitigation doctrine, which limits buyers' remedies by denying compensation for any losses that could have been avoided by reasonable mitigation.

As with many legal concepts, what counts as "reasonable" mitigation is a matter of some dispute. To the extent that courts' definition of reasonable mitigation corresponds with efficient mitigation, this limit on remedies can give buyers efficient incentives on this dimension.[85] Obviously, though, the efficacy of this mechanism depends on courts' ability to verify the costs and benefits of various mitigation activities.

5.2.6 The decision to terminate a project

In many cases, performance of a contract requires a sequence of many choices or events. For example, the construction of a building involves hundreds of separate steps whose performance can extend over weeks or months. Moreover, sometimes the earliest steps in the sequence must be taken at a time when it is still uncertain whether it will be worthwhile to perform all of the later steps. For example, if future construction costs are uncertain, it might (or might not) prove uneconomic to finish the building, but the early stages of construction may have to start before anyone can be certain what the eventual costs will be.

In such cases, each step in the sequence can be thought of as involving a choice between (a) continuing to perform, or (b) giving up the attempt to perform by terminating the project. Terminating the project early can allow the parties to take steps to reduce their losses, as discussed in the previous section on mitigation. On the other hand, continuing to perform preserves the option value of the project, by deferring to each later stage the decision to continue or terminate. The optimal decision, therefore, is to terminate early only when the savings from doing so exceed the project's option value (Triantis and Triantis, 1998).

The doctrine of anticipatory repudiation allows the potential breacher—the seller, in the examples above—to terminate a contract early simply by repudi-

[85] For discussions of this possibility, see Wittman (1981) and Goetz and Scott (1983).

ating it.[86] While this renders the seller liable for damages for breach, it also triggers the mitigation doctrine discussed above, thus requiring the buyer to begin taking any available steps to reduce those damages. As a result, if the seller faces a high probability of being unable to perform, she may be able to reduce her eventual liability by terminating the contract early. Viewed from this perspective, a termination decision by the seller can be thought of as just another form of breach, which will be efficient (or not) depending on whether early termination is optimal.

Indeed, if the seller can be held fully liable for all the losses from breach—either all the current losses, if the contract is terminated early, or all the subsequent losses if the contract is not terminated early—then she will have an incentive to make the optimal choice between continuing or terminating early, for much the same reasons discussed in the section on efficient breach (§5.2.1). However, holding the repudiating party liable for all losses may be difficult. The losses from an early termination include the loss of the option value of the contract, which may be difficult for courts to measure; omitting this loss from the measure of damages will bias the seller's incentives toward terminating too early (Triantis and Triantis, 1998). At the same time, the losses from any eventual non-performance may also be difficult to measure; or they may be difficult to collect, if the seller is by then insolvent. Omitting these losses from the damage measure will bias the seller toward terminating too late (Craswell, 1990).

Perhaps because of these difficulties, various legal doctrines also allow the non-breaching party—the buyer, in the example used here—to force an early termination (while still holding the other party liable for damages) subject to *ex post* judicial review. From the buyer's standpoint, a decision to terminate early can be thought of as another form of precaution (albeit an extreme one), which reduces the losses that will be suffered in the event of a breach. Equivalently, a decision not to terminate early can be thought of as another form of reliance by the buyer, since such a decision increases the gains if the project is eventually successful, but also increases the losses if the project eventually fails. In keeping with the earlier analysis of reliance and precautions (§5.2.3 and §5.2.4), the buyer will not have an incentive to make this decision optimally if he is fully insured by being guaranteed full compensation for his losses.

Perhaps for this reason, the law does not give the buyer the unrestrained power to elect an early termination of the contract, but (using various legal doctrines) subjects that decision to judicial review. For example, if the seller has already breached the contract during early stages of performance, the buyer can elect to force an early termination as long as those breaches are judged by a court to be "material" or "total." Alternatively, even if the seller has not yet violated any contractual requirement, the buyer may still elect to force an early termination if he has "reasonable grounds for insecurity" about the seller's performance, and if the seller is unable to provide "adequate assurances" that she will be able to perform.

[86] for discussions of the measurement of damages in cases involving anticipatory repudiation, see Jackson (1978) and Goetz and Scott (1983).

Of course, the exact effects of these doctrines depend on how courts interpret such vague standards as "material," "reasonable," or "adequate." (If courts could observe all the factors necessary to determine whether early termination was optimal, they could allow termination in exactly those cases; but their ability to do this is doubtful.) The issue is further complicated by the fact that courts do not usually issue advance opinions on pending disputes, but instead render their judgments in hindsight. An insecure buyer must thus decide whether to force an early termination before he is certain whether a court will agree that the other party's earlier breach was "material," or that the other party's assurances were not "adequate." If the buyer tries to terminate early, and a court later rules that grounds for early termination were lacking, the buyer may himself be held liable for breaching the contract prematurely. Given the substantial uncertainty associated with such vague legal standards, a risk-averse buyer may prefer not to force the issue (and a less risk-averse seller may take advantage of the situation to force a modification or a release of its obligation to perform).

Finally, as with other perform-or-breach decisions, the decision to terminate early should be made optimally whenever *ex post* renegotiation is possible, as there should always be some combination of side-payments that makes it in both parties' interests to terminate early whenever early termination is optimal. However, the legal doctrines discussed above will affect each party's share of any surplus, depending (for instance) on whether the buyer has a unilateral right to force an early termination without the seller's consent. As usual, though, such renegotiation may itself be costly; and the prospect of higher or lower side-payments (in the event of renegotiation) can also alter many of the other incentives discussed above, such as the incentive to rely or the incentive to take precautions.[87]

5.2.7 Decisions to gather and disclose information

Much of the literature described in the preceding subsections presupposes that both parties already know all of the relevant parameters, such as the distribution from which the seller's cost, c, will be drawn, or the extent of a buyer's reliance investments, r. In many cases, though, this (and other) information cannot be discovered without costly investments in information-gathering. To be sure, a number of legal doctrines affect the incentives to gather or disclose information, including some that are not normally classified as "remedies" doctrines (for example, see the discussions of fraud and mistake earlier in §2.5.2 and duty to disclose in §3.3.3). However, the remedies for breach can also affect various information-related incentives. While there are many different kinds of information that might be gathered or disclosed, and the effects of remedies on these incentives have not been studied as extensively as have the other incentives discussed earlier, we discuss some representative examples.

[87] For further discussion of all of these issues, see Goetz and Scott (1983), Craswell (1990), and Triantis and Triantis (1998).

Deciding whether to enter into contracts. The greater the penalties for breach, the greater will be sellers' incentive to gather information about potential risks that might leave them unable to perform. If a seller who fails to perform is held fully liable for all of a buyer's losses, that will make her internalize the full costs of any failure on her part to gather sufficient information. However, this will not be sufficient to make sellers' incentives optimal unless sellers can also internalize the full benefits of any additional investments in information-gathering. A perfect price-discriminating monopolist may be able to internalize all of those benefits, by charging higher prices; and in perfectly competitive markets, competitive pressures may force sellers to account for those benefits as well. In markets that fall between these extremes, however, it is more difficult to design a remedy that gives sellers the optimal incentive to gather this information (Craswell, 1988).

Searching for contractual partners. In many markets sellers (and buyers) may have a choice as to how much effort they devote to finding potential trading partners. Here, too, the parties' incentives depend partly on other legal doctrines—in this case, the rules governing initial contract formation (see §3 above). However, the parties' incentives on this point, too, will also depend on the general remedies for breach. For example, the steeper the penalties for breach, the harder it will be to get out of any contract once a contract has been formed, and so the more it will pay parties to invest in searching longer and harder to make sure they have found the best deal. Here, too, some of the benefits of increased search may be felt by the party who is found, rather than by the party who is doing the searching, so efficient search incentives can be created only if the searching party is able to internalize these benefits. For formal models incorporating these effects, see Diamond and Maskin (1979, 1981).

Optimal precaution and reliance. Even after one party has gathered information, in some cases the other party is the one who needs to be given that information in order to make an optimal decision. For example, if the efficient level of seller precautions depends on how much the buyer would gain from performance, and if the buyer already knows that information, the seller's precaution incentives can more easily be optimized if the buyer can communicate his information to the seller (Bebchuk and Shavell, 1991). Similarly, if the optimal level of reliance by the buyer depends in part on the probability that the seller will fail to perform, and if the seller already knows that probability, the buyer may be better able to choose an optimal level of reliance if the seller tells him what that probability is (Craswell, 1989b). While most of the conventional remedies for breach do not create any incentives for a buyer to convey this information to the seller, some remedies affect this incentive by conditioning the measure of damages on the information that has been disclosed. Under the rule of *Hadley v. Baxendale*, for example, a buyer facing large losses from breach is more likely to be allowed to recover those losses if she has told the seller about them in advance.[88]

5.2.8 The effects of party heterogeneity

Information also matters if sellers (or buyers) must deal with a heterogeneous population of trading partners. In many settings, for example, sellers may differ in the probability that they will be able to perform, or buyers may differ in the amount they would lose if the contract were breached. If each party's type is fully known to the other party, those differences can be fully reflected in each side's prices and other behavior, allowing each interaction to be analyzed as if it involved entirely homogeneous parties. When differences in characteristics cannot be fully observed, however, much of the analysis described above must be altered in some way.

For example, if buyers differ in the amount they will lose from a breach—and if the law's damage measure reflects these differences, by holding sellers liable for different damages depending on the buyer's type—sellers' expected liability costs will be lower when dealing with low-damage buyers than when dealing with high-damage buyers. If sellers can recognize the high-risk buyers, they can adjust their prices or their level of precautions or both to reflect those greater risks. However, this gives high-damage buyers an incentive to conceal their greater riskiness, if they can. In that case, the equilibrium will depend on the ability of the low-damage buyer to signal his type or the seller's ability to screen buyer types. As discussed in §2.3.2, signaling and screening can result in a loss of welfare. If screening or signaling are ineffective at separating the types, then a pooling equilibrium could result in which low-damage buyers effectively subsidize high-damage buyers. Alternatively, to avoid this subsidy, low-damage buyers could exit; that is, an adverse selection ("lemons") problem ensues. Limits on the recovery of unusually high damages—including the doctrine of *Hadley v. Baxendale*, discussed in the preceding subsection—could conceivably be used as a screen to prevent subsidization or drop out(see Quillen, 1988).

Buyer heterogeneity is also important if a seller has market power, insofar as it introduces the possibility of price discrimination. To be sure, if the seller can fully observe each buyer's type, she can engage in first-degree price discrimination, a practice that general contract law does not restrict. But if the seller cannot observe each buyer's type directly, she may be able to use some of the terms of her contract (including the remedies for breach) to separate different classes of buyers—screen—thereby increasing her profits through second-degree price discrimination (see, *e.g.,* Matthews and Moore, 1987). This possibility could also be relevant to the analysis of liquidated damage clauses, discussed later in §5.3.4. As discussed in §2.3.3, especially footnote 29, the welfare consequences of such price discrimination are *a priori* ambiguous.

[88] For economic analyses of this aspect of the Hadley doctrine see, *e.g.,* Ayres and Gertner (1989), Wolcher (1989), Johnston (1990), Bebchuk and Shavell (1991), Adler (1999), and Ben-Shahar and Bernstein (2000). Because of possible adverse price discrimination *vis-à-vis* the shipping price, however, a buyer may be reluctant to reveal that she faces large losses (see §4.2.4 and §5.2.8).

5.2.9 Risk allocation and insurance

Finally, in addition to all of the incentive effects discussed above, remedies for breach also have the effect of allocating various risks between the two parties. For example, the expectation remedy defined in §5.1.2 will, if it is measured perfectly, leave the buyer (the non-breaching party) fully insured. Thus, if buyers are risk averse, while sellers are risk neutral, this remedy will be efficient in terms of its affects on the parties' attitudes toward risk. For other combinations of attitudes toward risk, remedies other than expectation damages will be superior (see generally Polinsky, 1983). Unless one of the parties is actually risk loving, however, remedies that exceed the non-breacher's actual losses (*e.g.,* punitive damages) will almost always be undesirable as far as risk sharing is concerned (Craswell, 1996a).[89]

One case of interest concerns breaches that inflict non-pecuniary losses that do not increase the buyer's marginal utility of money (see Cook and Graham, 1977). For example, a photographer's failure to take wedding pictures might reduce the welfare of the marrying couple, but that does not mean they would prefer to buy an insurance policy that would pay them additional money if their wedding pictures were lost.[90] Viewed purely from the standpoint of the parties' taste for insurance, then, it might be better if contract remedies did not compensate for this sort of loss (as, indeed, they generally do not). At the same time, though, excluding these losses from contract remedies could also distort some of the other incentives discussed above, such as the photographer's incentive to take adequate precautions. For an analysis incorporating both of these effects, see Rea (1982).

5.3 Complications in determining monetary damages

In the discussion so far, we have dealt with the incentive and insurance effects induced by monetary remedies for breach. To be sure, much of the literature on this topic developed by analyzing particular legal remedies: either particular measures of monetary recovery, or non-monetary remedies such as specific performance. Reflecting that literature, §5.3.1 considers various legal and practical limits on contract damages; and section §5.3.2 considers some alternative damage measures that are imposed in certain cases. Then, §5.3.3 and §5.3.4 discuss specific performance (injunctive relief) and liquidated damage clauses (remedies stipulated in the contract itself).

[89]Risk aversion also plays a role in why limits on damages could be Pareto improving when the damage terms signal information about the potential breacher. See §2.3.2 and the discussion of Aghion and Hermalin (1990).

[90]Such preferences are, admittedly, inconsistent with neoclassical economic theory unless the couple exhibits lexicographic preferences in photos-money (other goods) space.

5.3.1 Limits on the measure of damages

In theory, expectation damages leaves the non-breacher just as well off in every state of the world (see §5.1.2); but in practice, a number of legal doctrines limit the losses that expectation damages will compensate. For instance, the non-breacher must prove the amount of his loss with "reasonable certainty"; often this will exclude the recovery of "speculative" losses whose amount was uncertain. Also, as noted earlier, contract law only rarely allows compensation for emotional losses (see §5.2.9). Perhaps more significant in this regard than any rule of contract law, however, is the general default rule of US legal procedure that requires each party to bear his own costs in litigation. In most other Western legal systems, in contrast, prevailing litigants are entitled to recover attorneys' fees as well as other out-of-pocket litigation costs. As a result, they come closer to being made whole than do winning litigants in the US.

By reducing the effective amount of the remedy, doctrines such as these weaken many of the seller's incentives discussed earlier (for example, the seller's incentive to take precautions, see §5.2.2). Of course, by shifting more of the loss to the buyer, the same doctrines may also strengthen the buyer's incentive to take various precautions (see, *e.g.,* §5.2.4). Also, when these exclusions are conditioned on particular behavior by the buyer—*e.g.,* if recovery for certain losses is excluded unless the possibility of such losses is disclosed in advance, under the doctrine of *Hadley v. Baxendale*—that could strengthen the buyer's incentives in more focused ways (see §5.2.7). Finally, if buyers differ in the extent to which they suffer non-recoverable losses, excluding those losses from the damage measure may reduce the cross-subsidization that could otherwise result (see §5.2.8).

5.3.2 Other measures of monetary damages

The damage measures identified above in section §5.1.2 above have received the most attention in the law and economics literature, but other measures are also sometimes used. As we show, though, most of the economic effects of these alternate measures can be decomposed into one or more of the effects already considered above.

Cost of completion. One recurring issue involves sellers who breach by leaving work unfinished, or by performing work work incorrectly, when it would be extremely costly to finish or correct the work. For example, suppose that the buyer would have realized net gains of v from complete performance, while leaving the work in its current state would leave the buyer with net gains of u, and it would now cost $f > v - u$ to finish the work as it should have been performed. If v and u have been measured correctly, this implies that it would be inefficient to finish the work. However, courts vary in their treatment of this case, sometimes allowing the buyer to recover the completion cost f in damages, while sometimes limiting the buyer to recovery of $v - u$.

Awarding f in damages would be inefficient if it led to the work actually

being completed. If, however, completion would in fact be inefficient, then the buyer will pocket the damage payment rather than use it to finish the work. The principal effects of awarding f, rather than awarding $v-u$, will, thus, be those discussed previously. The larger remedy will (in general) strengthen sellers' incentives to avoid committing this sort of breach; perhaps strengthening them excessively, if the actual losses caused by the breach are only $v-u$. On the other hand, if $v-u$ actually understates the buyer's losses—say, because some of the benefits from performance are hard to measure, and have therefore been omitted from the court's measure of v—then increasing the remedy to f could improve the seller's incentives in some respects (see Muris, 1983, for a discussion of both effects). Even then, much would depend on the exact nature of the benefits that were excluded, as was also discussed in preceding subsections. For example, if buyers differ in the extent to which they would realize certain benefits, excluding those benefits from the damage measure could reduce any cross-subsidization that might otherwise result (see §5.2.8).

DISGORGEMENT. Similarly, courts occasionally require a breaching party to disgorge any profits he may have earned as a result of the breach, even if those profits exceed the non-breacher's expectation loss, often citing a general principle that no one should profit from his own wrong. This result is not the norm in contract cases, but is reserved for situations in which the breacher is thought to have engaged in bad faith or "willful" breach (an ill-defined notion in the case law) or when the non-breacher is considered to have a property or quasi-property interest in the subject of the contract and is thus entitled to the proceeds of resale, even if he could not have earned such proceeds on his own (see Farnsworth, 1985, for a general discussion).

Disgorgement damages, if assessed with certainty, leave the breaching party indifferent between performance and breach. As such, they entirely eliminate any incentive to breach, which from the viewpoint of efficient breach, is an excessive deterrent. In turn, the absence of efficient breach could also generate excessive reliance by the promisee.

A rationale sometimes articulated in favor of disgorgement damages is that, by removing any incentive for unilateral breach, they encourage a party who would like to escape performance to approach the counterparty and negotiate a modification or release (see, *e.g.,* Friedmann, 1989). Such negotiation may be desirable if the potential breacher would otherwise be uninformed about the size of the counterparty's expectation loss (thus leading him mistakenly to breach), or if the transaction costs of renegotiation are less than the costs that would be occasioned by a lawsuit. The frequent association of disgorgement with the elements of bad faith or willfulness, however, suggests an underlying punitive component to this remedy, which raises the topic of punitive damages generally.

PUNITIVE DAMAGES. In general, explicitly punitive damages are rarely imposed in contracts cases. This is in part because breach of contract, though a legal wrong, is typically judged more leniently than would breaches of duty in a tort

or property setting, for a number of related reasons. First, because liability is based on a voluntary exchange relationship, the expected costs of breach and of paying damages are likely to be reflected in the contract price, so that efficient breaches work to the *ex ante* advantage of both parties. Second, to the extent that punitive damages are justified by negative externalities imposed on the general community, the concern is less pressing in the ordinary contract. Third, to the extent that punitive damages are justified by a high likelihood of non-detection, this concern is lessened when the parties know each other and are likely to be watchful of proper performance.

Punitive damages are very occasionally imposed, however, in response to breaches of contract that also involve a tort, gross unfairness, or a violation of some public policy. For example, punitive damages have regularly been imposed on insurance companies that refuse without justification to pay valid claims. This result is often defended in terms of the imbalance in the parties' economic power, the particularly difficult circumstances in which such refusal places the insured, and the likelihood that most victims of such opportunism are likely never to seek redress in court—all factors that go to the general case for punitive damages, as laid out in succeeding chapters of this handbook.

5.3.3 Specific performance

In some cases courts order specific performance rather than awarding monetary damages for breach. In effect, this remedy requires the seller (or other breaching party) to perform the contract in full, backed by the threat of fines or even imprisonment if she fails to do so.

If *ex post* renegotiation is impossible, such an order would lead to inefficient performance of the contract in any case where breach was more efficient. As long as the parties can renegotiate, however, they should always be able to avoid this loss by agreeing not to perform, with the gains from non-performance being shared between the parties in accordance with their bargaining strength. But if the buyer has the threat of a remedy of specific performance, thereby requiring the seller to incur the costs of performance, that should allow the buyer to capture more of the gains than he could if his only legal threat were to hold the seller responsible for some smaller monetary remedy.

As a result, when *ex post* renegotiation is possible, the effects of specific performance will be felt in all of the ways discussed above. When renegotiation is costly, specific performance could, in principle, add to those negotiation costs, though it is unclear whether the negotiations required under specific performance will be any more or less costly than those required under any alternative remedy.[91]

[91] For analyses of this issue see, *e.g.,* Kronman (1978b), Schwartz (1979), Ulen (1984), and Bishop (1985). Most recently, Shavell (2006) has argued that negotiation costs will be relatively large where the reason for breach is a production cost increase (because in order to negotiate a release, the parties must agree on how to distribute the quasi-rents arising from the cost increase), and resulting), and relatively small where the reason for breach is to sell to an third party (because the buyer can also resell to the third party, and so the quasi-rents

In any event, if specific performance results in the buyer being able to negotiate for a larger payment, this will have the same effect as an increase in the expected size of any monetary remedy. For example, the threat of having to make such a payment will strengthen the seller's incentives to take precautions against events that might expose it to such a remedy (Muris, 1982). Moreover, this threat will also increase the buyer's incentives to make reliance investments (Edlin and Reichelstein, 1996). An increase in the expected payment will also alter the risks born by each party, as discussed earlier in §5.2.9. In general, specific performance makes it more likely that the buyer will end up at least as well off as if the contract had been performed (otherwise, he would not consent to any renegotiation), whereas the award of smaller monetary remedies might not leave the buyer this well off. But whether this increase in compensation is desirable, all things considered, depends on all of the effects discussed earlier, as compared to the effects produced by whatever monetary measure of damages the court would award if it did not require specific performance. In this regard, the choice between specific performance and monetary damages has much in common with the choice between injunctive relief and monetary remedies in many other areas of law.[92]

Under common law, specific performance has traditionally been more difficult to obtain than monetary damages, with injunctive relief treated as matter of the court's discretion, and usually being reserved for cases in which damages are deemed insufficient to protect the non-breacher's expectation interest or in other special circumstances (for example, when the goods being traded are unique, when the breacher is insolvent, or when the non-breacher has made relational investments that would be difficult to replace). The most widely cited policy reason for these restrictions is that specific performance is thought to impose greater administrative costs on the legal system, especially in situations where the quality of a coerced performance is costly to verify. But this concern has not prevented civil law systems from making specific performance their remedy of default, even though in many circumstances their courts will still award money damages in substitution for performance (Lando and Rose, 2004).

5.3.4 Remedies expressly stipulated in the contract

As noted earlier (see §5.1.2), the remedies discussed above are usually supplied by the legal system as default remedies to be applied in cases where the contract is silent as to the consequences of breach. In some cases, though, the parties' contract may itself stipulate the remedy that will be required in certain states of the world. Usually this remedy will consist of a monetary payment, but in some cases, the parties may provide for particular actions to be taken, as when a sales contract provides that in the event of a defect in the goods, the seller

are limited to the difference between the buyer's and seller's cost of resale).

[92] There is a much more extensive literature analyzing the analogous issue in property and tort law. See, *e.g.,* Calabresi and Melamed (1972), Kaplow and Shavell (1996), Bebchuk (2001), and Ayres and Goldbart (2003).

will be obliged to provide repair or replacement. When such stipulations take a monetary form, they are usually referred to as "stipulated damages" or, at least when they are enforceable, "liquidated damage clauses."[93]

Depending on the amount chosen by the parties to serve as the measure of damages, liquidated damage clauses can produce any and all of the effects described above.[94] For example, clauses that specify a large payment can give sellers a strong incentive to avoid committing a breach, while clauses that specify smaller payments will give buyers less insurance against breaches. The amount of the clause can also affect each side's incentives to rely on the contract, or to take precautions against various risks, as the preceding sections also discussed. Indeed, while the literature summarized in the preceding sections was mostly written to provide guidance to courts or other lawmakers, that literature can just as easily be read as providing guidance to private parties who wish to select an efficient liquidated damage remedy (see Katz, 1996b).

Moreover, in some respects liquidated damage clauses (drafted by the parties) are likely to be superior to general default rules (selected by courts or legislatures). As the preceding sections make clear, most remedies involve trade-offs among various important incentives, and in many cases the contracting parties are better suited than courts to choose the particular trade-off that is best for their own transaction. In markets where parties are heterogeneous (see §5.2.8), liquidated damage clauses can be tailored to particular contracting pairs. More broadly, all of the reasons that support the enforcement of contracts generally (see §2.2) will usually argue for the enforcement of liquidated damage clauses in particular.

It is therefore striking that common-law courts refuse to enforce clauses that set damage amounts that the courts consider excessive (these are typically referred to as "penalty clauses"). This reluctance may be due partly to historical factors, and in particular to the belief (on non-economic grounds) that "penalties" or "enforcement" should be the exclusive province of the legal system, rather than being subject to private control. In addition, though, there may also be economic reasons that—in particular situations—might counsel against enforcement.

For example, in cases where the actual damages from breach turn out to differ significantly from what the parties expected, it is possible—though far from automatic—that a remedy specified by the liquidated damage clause might no longer be welfare maximizing.[95] In effect, denying enforcement in these cases would be similar to denying enforcement of the contract itself, under the "im-

[93] Very occasionally, parties will provide for specific performance in their contracts, but courts do not regard themselves as bound by such provisions (although they will take them into account as a factor bearing on their exercise of discretion).

[94] For discussions of these effects in connection with liquidated damage clauses, see Goetz and Scott (1977), Clarkson et al. (1978), Rea (1984), Schwartz (1990), and Edlin and Schwartz (2003).

[95] Observe, however, that if the remedy is a monetary transfer, then welfare is typically assumed to be unaffected by transfers.

plied excuse" doctrines of mistake or impracticability, in cases where unforeseen events have significantly altered the contracting parties' situation. It is, however, a matter of debate whether courts have the ability to make such *ex post* adjustments in ways that will in fact improve the parties' *ex ante* welfare. These doctrines are discussed above in §4.4; for discussions focusing specifically on liquidated damage clauses, see Rea (1984) and Schwartz (1990).

In addition, some liquidated damage clauses can affect the welfare of others who are not parties to the contract. In particular, if a seller with market power is concerned with defending its market against entrants, liquidated damage clauses can serve as a commitment to deter competitors from entering (Aghion and Bolton, 1987; Chung, 1991; Spier and Whinston, 1995). Even in competitive markets, if some or all buyers are unaware of seller's liquidated damage clauses, this imperfect information could create incentives for socially undesirable clauses (just as they could in the case of any contractual clause, see generally §2.3.2).

It is worth noting that there are some situations in which common-law courts are more likely to enforce privately stipulated remedies. In particular, courts are more deferential to liquidated damage clauses that turn out to be undercompensatory *ex post*, as compared to those that turn out to be over-compensatory (perhaps because the latter raise third-party effects where the former do not). They are quite deferential to clauses that disclaim liability for consequential damages and that limit the remedy for breach of warranty to repair or replacement (because such clauses supplement the doctrine of *Hadley v. Baxendale*, of which the courts approve). They are also likelier to enforce liquidated damages that have actually been paid over as an advance deposit, although even in those cases, the breacher may be entitled to restitution of part or all of the deposit to the extent it plainly exceeds the non-breacher's expectation loss.

5.4 Private enforcement of contracts

As noted, legal enforcement or its threat are not the only means by which parties are induced to honor their commitments. In this subsection, we briefly consider some of these other means. In §5.4.1, we briefly consider how the costs of using legal enforcement can either distort contracts or cause the parties to dispose of them altogether. In §5.4.2, we take up how repeated interactions or reputational concerns can deter breach. Finally, §5.4.3 discusses various legal doctrines that bear on, and in some cases support, these private alternatives to traditional legal enforcement.

5.4.1 Enforcement costs

One reason contracts could fail to be enforced as written is that enforcement requires expenditures by the parties that are either *ex post* incredible or can be anticipated to be so large *ex ante* that no contract is written. A partial list of examples is: (1) the agreement is illegal or exceeds the parties' power to contract under the applicable legal system (see §2.5 above); (2) courts cannot verify critical aspects of contractual performance or whether relevant contingencies have arisen (see §4.2 above); (3) litigation is costly in terms of time, risk, or material resources; (4) the defendant may be insolvent or otherwise lack the ability to comply with a judgment; and (5) the dispute arises in the course of an otherwise successful relationship that the parties do not wish to jeopardize. In addition, as a large literature makes clear, two additional reasons are also important in developing countries:[96] (6) the court system operates corruptly; and (7) courts are incapable of enforcing their verdicts, because police are corrupt or unavailable.

A simple model illustrates some of these issues. Suppose that party B employs party A and promises to pay A w upon completion of A's task. Suppose it costs A $k_A > 0$ to have the contract enforced or fight litigation and it costs B $k_B > 0$ for the same. If A cannot recover its enforcement costs from B should she prevail at trial, then B knows he can underpay A by up to k_A without A seeking to enforce the contract (assuming A acts in a coldly rational way).[97] If, however, the parties anticipate this, then they could agree to a nominal wage of $w + k_A$, recognizing that B will underpay by k_A. Somewhat more problematic is non-performance by A. If it is feasible for A to overperform by an amount k_B, then the parties can simply set the performance standard in the contract k_B above what they truly intend. But such overperformance could be infeasible, in which case this trick won't work. Now it could be necessary to add a clause to the contract that A pay B k_B should A be found to have underperformed. To the extent the court refuses to honor that clause, citing the unenforceability of penalty clauses (see §5.3.4), or its inability to verify A's performance adequately, this solution could also fail. If it is impossible to enforce a contract contingent on A's performance, then the parties will either have to forgo contracting or they will have to contract around the problem (*e.g.,* use a revenue-sharing contract to give A better incentives).

More generally, whenever judicial enforcement is likely to occur in less than 100% of the cases, it might in principle be possible to make up for that deficiency by increasing the size of the damage award in those cases that do reach the courts. To take a simple example, if only one out of ten breaches is ever sanctioned by courts, many of the incentive effects could be restored if the damage award that would otherwise be optimal were multiplied by ten in every case.[98] However, courts are usually reluctant to make this adjustment in contract cases, where punitive damages are only rarely awarded. This solution will also be unattractive if either party is risk averse, as it increases the variance of both parties' returns (see §5.2.9). And in more extreme settings, where litigation is infinitely costly (*e.g.,* performance is completely unverifiable or in countries

[96] See, for instance, Anderson and Young (2002), Cungu and Swinnen (2003), or, for a more historical perspective, Greif and Kandel (1995).

[97] It is known from research on ultimatum games (*e.g.,* Güth et al., 1982) that players do not always act in a coldly rational way, preferring sometimes to punish others even if the act of punishing is costly to them. See Rabin (1993) for a discussion and analysis.

[98] This effect is often suggested as an economic rationale for punitive damages. See, *e.g.,* Polinsky and Shavell (1998) and Craswell (1999).

where a reliable court system is unavailable), reliance on court-ordered remedies is bound to fail.

Unfortunately, the issue of credible enforcement is typically ignored in most of the contract design literature.[99] Accordingly, we now turn briefly to enforcement methods that do not requires the participation of courts.

5.4.2 Self-enforcing contracts

It has long been understood from the repeated games literature that some agreements are self enforcing in the context of an ongoing relationship.[100] The most prominent example of such "agreements" is tacit collusion among competing firms. That is, recognizing their repeated interaction, firms avoid undercutting each other on price. This "agreement" to keep prices high is enforced by the threat of a price war should any firm undercut.[101]

Within the realm of contracts, there is wide scope for such self-enforcing agreements. Moreover, self-enforcement can substitute for legal enforcement. For instance, in a one-shot game, an employer might choose to renege on a promised wage payment to an employee (recall the discussion in the previous subsection). But in a repeated context, the employee could retaliate by quitting (or possibly engaging in sabotage), which could be sufficiently costly to the employer that he chooses to pay the employee.[102] Of course, the employee could also take the employer to court for nonpayment, but, as we saw above, legal enforcement might not always be credible.

Self enforcement can also complement legal enforcement. For instance, while it might not be credible for a party to enforce a contract legally in a one-shot game, the party might wish to develop a reputation for enforcement in a repeated context—just as in some models of entry deterrence whereby an incumbent firm punishes entrants to establish a reputation for toughness (see, *e.g.,* Tirole, 1988, Chapter 8), a party may want to develop a reputation as someone who can't be cheated (*e.g.,* become known as litigious). Of course, if repetition is what is making legal enforcement of contracts credible, then one should model repetition explicitly in the analysis of the underlying contracting problem if that problem is itself repeated (*i.e.,* the game between an employer and a long-term employee). In other words, an appeal to reputation for enforcement in a static analysis of a contracting problem is most acceptable when the problem is short term (*i.e.,* a given pair meet only once to contract), while the players are long lived (*i.e.,* will play again with others).

Within the literature, self-enforcing contracts are often known as *relational contracts*. Applications have included quality assurance for experience goods

[99] Some notable exceptions are Spier (1994) and Krasa and Villamil (2000).

[100] For a review of repeated games see Chapter 2 of Gibbons (1992) or Chapter 5 of Fudenberg and Tirole (1991).

[101] See Chapter 6 of Tirole (1988) for details.

[102] See, also, Thomas and Worrall (1988).

(*i.e.,* goods the quality of which can only be assessed via consumption);[103] incentive schemes;[104] and social contracts within firms.[105] However, explicit models of reputation in the legal literature on contacts are still relatively rare.[106]

In addition, while reputations can, in many markets, provide powerful incentives to perform, in some markets they are less likely to be effective. For example, if performance involves a credence good (the quality of which cannot be observed even after consumption, at least not without expert diagnosis), many breaches may go undetected, with little harm to the breaching party's reputation.[107] In addition, the enforcement of reputations may be privately costly to those who enforce them, thus leading to free-rider problems in enforcement.[108] In other markets, where sellers' histories are not easily discoverable by buyers (or *vice versa*), the incentive effects of reputations may be weakened, though the involvement of kin or ethnic groups or other reputational intermediaries may help in overcoming that difficulty.[109] Finally, for sellers who are on the verge of bankruptcy (or are otherwise reaching their "last period"), the prospect of losing future business may be a very weak constraint at best.

5.4.3 Legal doctrines affecting self-enforcement

As the preceding subsection discussed, reputations are most effective in the context of a repeated game, so that a party who cheats suffers the consequence by losing the benefit of future interactions. In some contractual settings, however, the party who cheats can be made to suffer an extra-legal sanction in connection with the very contract has that been breached, as long as the other party has not yet fully performed his own end of the contract. In such a case, the other party may respond not by filing a lawsuit, but by withholding his own performance. For example, if a seller agrees to deliver goods on credit, but if the buyer discovers (before paying) that the goods are defective, the buyer might respond to this breach by refusing to accept or to pay for the goods.

To be sure, the significance of a threat to suspend performance depends partly on the value that performance has to the other party, but it also depends on how the parties have structured their transaction. To take an extreme case,

[103] Klein and Leffler (1981) and Shapiro (1983) are two examples.

[104] Bull (1987) and Levin (2003) are two examples.

[105] See Hermalin (2001) for a survey.

[106] For a qualitative discussion, see Charny (1990). Bernstein (1992, 1996, 2001) has presented several case studies illustrating the operation of reputational enforcement in specialized markets.

[107] See Darby and Karni (1973). Although see Fong (2005) for a more nuanced analysis.

[108] For example, Klein and Leffler (1981) model a market in which buyers, if they even once receive a defective product, follow a flat rule of never purchasing from that seller again. While such a rule does produce desirable incentive effects, it may or may not be rational for individual buyers.

[109] See Landa (1981) and Bernstein (1992) for discussion of kin and ethnic networks. See Mann (1999) on other reputational intermediaries.

if the contract calls for the seller to make all her deliveries before the buyer pays any of the price, that gives the buyer a good deal of leverage by threatening to withhold payment. On the other hand, if the contract instead calls for the buyer to pay the entire price in advance, before any of the goods have been shipped, this gives the buyer very little leverage, while putting the seller in the happy position of being able to threaten to suspend all of her shipments. As a consequence, parties often negotiate extensively over the exact timing of the various payment and delivery requirements.[110]

In addition, a party's right to suspend his or her own performance may also be regulated by various legal doctrines, as we now discuss.

Rescission. Once one party has committed a breach, the other party may sometimes be able to choose between monetary remedies (typically expectation damages, as discussed earlier in section §5.1.2) and simply walking away from the contract, without collecting any remedy at all. This latter option is usually referred to as termination or rescission. To be sure, if the contract would have been a profitable one for the non-breaching party, that party will usually prefer expectation damages over rescission, for expectation damages should give the non-breacher all of the benefits she would have received from performance (if all of those benefits can be adequately measured, see section §5.2.1 above). But if the contract in question would have been a losing one for the non-breaching party, rescission may be a more valuable remedy, as it allows that party to walk away from what might otherwise be a significant loss.

To complicate matters further, in some cases the non-breaching party may elect to rescind a contract even if she has already performed some part of her own services under the contract. By rescinding the contract, the non-breaching party would give up her right to recover the payment specified in the contract, but she could then sue for restitution to recover a judicially-determined "reasonable value" for her services.[111] As the court's determination of reasonable value need not be limited by the contract price for those services, this remedy could leave the non-breacher with more than she could get under any alternative remedy. Indeed, precisely because rescission is elective for the non-breacher, a rational non-breacher will not choose it unless it is more favorable to her.

However, several legal doctrines limit the use of rescission as a response to the other party's breach. At common law, breach of a service contract allows the non-breaching party to elect rescission only if the breach is "material" or "substantial," a vague test that leaves much to the courts' discretion. By contrast, breach of a contract for the sale of goods is said to allow rescission for any breach whatsoever (the so-called "perfect tender" rule). The Uniform Commercial Code has altered this latter rule, though, by limiting the buyer's

[110] See Scott and Triantis (2006). In addition to the sources cited earlier in §5.2.6, brief discussions can also be found in Goetz and Scott (1983), Kull (1991), and Kraus (1994). Note, also, the connection between this and the discussion of option contracts and holdup in §4.3.2.

[111] For other uses of restitution as a remedy, see section §5.1.2 *supra*.

right to rescind in cases involving the sale of goods. Under the UCC, defective goods in a single shipment of a multi-shipment (or installment) contract do not allow the buyer to rescind the entire contract unless the defect "substantially impairs" the value of the entire contract. Even in contracts calling for only a single installment, the buyer may lose the right to rescind if he fails to reject the installment within a "reasonable" time for inspection. The buyer's right to rescind may also be limited by the seller's right to take a "reasonable" amount of time to "cure" the defect (see the discussion of "cure" below).

While the remedy of rescission has not been analyzed as extensively as other remedies have, many of the effects are similar to those of any other remedy that is more generous to the non-breacher.[112] That is, as noted, the non-breacher will elect rescission only when it is more favorable to her than the other available remedies. Consequently, the availability of rescission should increase the breacher's incentive to perform—just as would any other increase in the size of the likely remedy. Of course, the parties may be able to renegotiate *ex post* to avoid inefficient breach or inefficient performance—but, as with other remedies, the payments that parties must make in *ex post* renegotiations will still affect their *ex ante* investment incentives. Also, if either party is risk averse, the availability of rescission will also affect their risk-bearing costs, again in the same manner as any other increase in the expected size of the remedy.

Cure. As noted in the preceding paragraphs, the buyer's right to rescind a contract for the sale of goods may be limited (under the UCC) by the seller's right to cure any defects in the goods that she delivered. As long as the time specified for delivery of the goods has not yet expired, the seller has complete freedom to try to cure the defects and deliver conforming goods. However, even after the time for delivery has expired, the seller may still have some right to attempt a cure, although this right is subject to various legal limits (many of which are vague). For example, the seller may not take more than a "reasonable" time to effect such a cure; and may only do so when she had "reasonable grounds" to believe that her original, non-conforming delivery would have been acceptable "with or without a monetary allowance" for the defect. In installment contracts, where only the goods in a single shipment were defective, the seller must be able to provide "adequate assurances" that its cure will be successful. And in all cases, it is ultimately up to a court to determine whether the seller's efforts have in fact cured the defect.

In cases where it is clear that the seller can fix the defect, the right to cure serves to limit the effect of the remedy of rescission. For example, if the market price of goods fell significantly after the contract was signed, the buyer might otherwise use a trivial defect—one that could be cured at a *de minimis* cost—to rescind the contract, thus forcing the seller to bear the loss from the market's fall. The economic effects of this use of rescission were discussed above. Clearly, giving the seller a right to cure eliminates those effects.

[112] The earliest economic discussion is Goetz and Scott (1983). Other discussions include Kull (1991) and Kraus (1994).

In other cases, though, it may not be clear (at least initially) whether the seller will ever be successful in curing the defect. For example, many litigated cases involve the sale of cars or houses that seem to be "lemons," whose seller makes repeated but unsuccessful attempts to find and fix the problem.[113] In these cases, a court's interpretation of the right to cure has the effect of determining the point at which the contractual endeavor should be terminated in order to cut the parties' losses. The economic implications of this decision were discussed earlier in §5.2.6.

Conditions and termination clauses. Parties can also use the contract itself to specify (within limits) the conditions under which either or both parties will be released from their obligation to perform. Employment contracts, for example, may allow either the employer or the employee to terminate the relationship at any time, and for any reason. (Indeed, this is the common-law default rule for employment contracts.) Similarly, franchise contracts may specify that the franchise relationship will continue indefinitely unless one or the other party exercises its right to terminate the relationship, often with some advance notice required (*e.g.,* 30 days' notice of termination).

Other contracts may permit one part to terminate the relationship if certain terms of the contract are violated. Technically, contractual clauses whose violation will release one party from part or all of the contract are referred to in law as conditions. By contrast, covenants or promises are clauses whose violation normally leads to some other default remedy. Violation of a covenant will release the other party from the contract only if the breach is found to be "material" (see the preceding discussion of rescission).

To be sure, just as courts sometimes refuse to enforce liquidated damage clauses (see §5.3.4), they also do not always enforce contractual termination provisions. For example, if termination would inflict on the other party a loss that seems to the court to be excessive (a "forfeiture"), courts may refuse to give effect to an express condition, thus prohibiting the other party from terminating the contract. In addition, clauses that purport to give one party the right to terminate a relationship for any reason will sometimes be interpreted more narrowly by courts, who may refuse to permit terminations that are not made "in good faith" (a phrase whose exact content is difficult to pin down). By requiring some degree of judicial approval of a termination decision, these doctrines thus limit the parties' ability to use termination as a self-enforcement technique. Some of the economic effects of these limits are discussed in Klein (1980).

[113] In some cases, the contract may itself specify that the seller has the right to cure, and may even limit the buyer's remedy to accepting the seller's repair or replacement. Even in these cases, though, the UCC allows courts to disregard such a clause (and bring the seller's right to cure to an end) if the seller's inability to cure causes this remedy to "fail its essential purpose."

5.5 Other law bearing on contract enforcement

It should also be kept in mind that the enforcement of contracts is often affected by rules and institutions from other fields of law. For example, one common way for parties to enhance the likelihood of contractual performance is to offer collateral; and this device is regulated generally by the law of property and specifically by the specialized law of secured transactions (see Schwartz, 1989; Triantis, 1992). Similarly, contracting parties often enlist third parties as guarantors on their behalf; and the value of such assurance is determined by the law of suretyship (see Katz, 1999).

Conversely, parties' ability to use reputation or repeat dealing as private enforcement devices may be restricted by other fields of law. For example, the law of antitrust generally prohibits concerted boycotts or refusals to deal; and the law of torts may treat some reputation-affecting communications as unfair competition, defamation, or invasion of privacy. To this extent, some private attempts to enforce contracts may be actionable in their own right.

6 Conclusions

This chapter is lengthy and many of our conclusions have already been given. Consequently, we limit ourselves here to a few remarks.

Given the vastness of the literature on the law and economics of contracts, even a survey as long as ours must omit certain topics. One topic that has been omitted is the connection between the literature on contracts and those on torts, takings, and regulation. In particular, much of the economic analysis of regulation takes the view that the regulator and the regulated entity are entering into what is, effectively, an agency contract between the regulator (the principal) and the regulated entity (the agent). See, for instance, Laffont and Tirole (1993). But even torts and takings can be related to contracts if, as some analysis has done, one views the state as seeking to approximate, in some way via law, the contract that it would have liked to have written with the tort malfeasor or the owner of the property to be taken if their identity were known in advance.[114]

We have also omitted the entire literature in which the state itself is a party to the contract. Government contracts raises a number of additional issues, including the need for public accountability, risks of corruption and political capture, the problem of establishing credible commitment that will survive changes of governmental regime, and the special difficulties of enforcing contract rights against a sovereign state. For instance, the risk of nations repudiating their debt contracts or abrogating licensing agreements is well documented, especially in the context of developing economies. Creditors or technology providers who contemplate entering into agreements with such governments, accordingly, must find ways to mitigate or insure against such risk. At the same time, citizens and regulators have an interest in preventing state officials from entering

[114] See, *e.g.,* Hermalin (1995) for an application of this approach on the takings issue.

into contracts that are not in the public interest *ex ante*. To an extent, these issues relate to our discussion of alternative means of enforcement in §5.4.2 and §5.4.3, but the area is broader than this.

Any chapter of this sort should close with some suggestions for future research. Some suggestions have made already: (i) more economic analysis of non-Anglo-American contract law; (ii) more positive analyses of contract law; (iii) more efforts in modeling to treat monitoring and measurement as endogenous with respect to what information is observable or verifiable; (iv) empirical studies of how courts employ certain rules, such as the parol evidence rule; (v) economic analyses of some of the doctrinal complications associated with the law of contract formation (*e.g.,* promissory estoppel); and (vi) more analysis of the interactions between private and state enforcement of contracts. To this list we would add greater use of new economic paradigms such as behavioral economics. The behavioral paradigm in particular holds out the promise of increased understanding of the phenomenon of bounded rationality, and of legal doctrines that respond to it. Consider, for instance, the literature discussed in §2.3.4 above, as well as more recent work such as that of DellaVigna and Malmendier on issues of contract design and self control and their application to questions of how do health clubs design their contracts (DellaVigna and Malmendier, 2004, in press, respectively).

References

Adler, Barry E., "The Questionable Ascent of Hadley v. Baxendale," *Stanford Law Review*, 1999, *51* (6), 1547–1589.

Aghion, Philippe and Benjamin E. Hermalin, "Legal Restrictions on Private Contracts Can Enhance Efficiency," *Journal of Law, Economics, and Organization*, Fall 1990, *6* (2), 381–409.

_ and Patrick Bolton, "Contracts as a Barrier to Entry," *American Economic Review*, June 1987, *77* (3), 388–401.

_ , Mathias Dewatripont, and Patrick Rey, "Renegotiation Design with Unverifiable Information," *Econometrica*, 1994, *62* (2), 257–282.

Akerlof, George, "The Market for 'Lemons': Qualitative Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Anderlini, Luca and Leonardo Felli, "Incomplete Written Contracts: Undescribable States of Nature," *Quarterly Journal of Economics*, November 1994, *109* (4), 1085–1124.

_ , _ , and Andrew Postlewaite, "Should Courts Always Enforce What Contracting Parties Write?," Working Paper, University of Pennsylvania November 2003.

Anderson, James and Leslie Young, "Imperfect Contract Enforcement," Technical Report 8847, NBER, Cambridge, MA 2002.

Arrow, Kenneth J., *Social Choice and Individual Values*, 2nd ed., New Haven, CT: Yale University Press, 1963.

_ and Gerard Debreu, "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, 1954, *22* (3), 265–290.

Atiyah, P.S., *An Introduction to the Law of Contract*, 6th ed., Oxford: Oxford University Press, 2003.

Ayres, Ian, "Valuing Modern Contract Scholarship," *Yale Law Journal*, January 2003, *112* (4), 881–902.

_ and Eric Rasmusen, "Mutual and Unilateral Mistake in Contract Law," *Journal of Legal Studies*, 1993, *22* (2), 309–343.

_ and Gregory Klass, *Insincere Promises: The Law of Misrepresented Intent*, New Haven: Yale University Press, 2005.

_ and Paul M. Goldbart, "Correlated Values in the Theory of Property and Liability Rules," *Journal of Legal Studies*, January 2003, *32* (1), 121–151.

_ and Robert H. Gertner, "Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules," *Yale Law Journal*, 1989, *99* (2), 87–130.

_ **and** _ , "Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules," *Yale Law Journal*, 1992, *101* (4), 729.

Baird, Douglas and Robert Weisberg, "Rules, Standards, and the Battle of the Forms: A Reassessment of 2–207," *Virginia Law Review*, 1982, *68* (6), 1217–1262.

Bar-Gill, Oren, "Seduction by Plastic," *Northwestern University Law Review*, 2005, *98* (4), 1373–1434.

_ **and Omri Ben-Shahar**, "The Law of Duress and the Economics of Credible Threats," *Journal of Legal Studies*, 2004, *33* (2), 391–430.

Barnett, Randy E., "A Consent Theory of Contract," *Columbia Law Review*, 1986, *86* (2), 269–321.

Barton, John H., "The Economic Basis of Damages for Breach of Contract," *Journal of Legal Studies*, June 1972, *1* (2), 277–304.

Bebchuk, Lucian Arye, "Property Rights and Liability Rules: The *Ex Ante* View of the Cathedral," *Michigan Law Review*, December 2001, *100* (3), 601–639.

_ **and Omri Ben-Shahar**, "Precontractual Reliance," *Journal of Legal Studies*, 2001, *30* (2, pt. 1), 423–457.

_ **and Steven Shavell**, "Information and the Scope of Liability for Breach of Contract: The Rule of *Hadley v. Baxendale*," *Journal of Law, Economics, and Organization*, Fall 1991, *7* (2), 284–312.

Ben-Shahar, Omri, "The Tentative Case Against Flexibility in Commercial Law," *University of Chicago Law Review*, 1999, *66* (3), 781–820.

_ , "Agreeing to Disagree: Filling Gaps in Deliberately Incomplete Contracts," *Wisconsin Law Review*, 2004, (2), 389–428.

_ , "An Ex-Ante View of the Battle of the Forms: Inducing Parties to Draft Reasonable Terms," *International Review of Law and Economics*, in press.

_ **and Lisa Bernstein**, "The Secrecy Interest in Contract Law," *Yale Law Journal*, 2000, *109* (8), 1885–1925.

Bernheim, B. Douglas and Michael D. Whinston, "Incomplete Contracts and Strategic Ambiguity," *American Economic Review*, 1998, *88* (4), 902–932.

Bernstein, Lisa, "Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry," *Journal of Legal Studies*, January 1992, *21* (1), 115–157.

_ , "Merchant Law in a Merchant Court: Rethinking the Code's Search for Immanent Business Norms," *University of Pennsylvania Law Review*, 1996, *144* (5), 1765–1821.

_ , "The Questionable Empirical Basis of Article 2's Incorporation Strategy: A Preliminary Study," *University of Chicago Law Review*, 1999, *66* (3), 710–780.

_ , "Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions," *Michigan Law Review*, 2001, *99* (7), 1724–1790.

Birmingham, Robert L., "Breach of Contract, Damage Measures, and Economic Efficiency," *Rutgers Law Review*, Winter 1970, *24* (2), 273–292.

Bishop, William, "The Choice of Remedy for Breach of Contract," *Journal of Legal Studies*, June 1985, *14* (2), 299–320.

Blackorby, Charles and David Donaldson, "Cash versus Kind, Self-selection, and Efficient Transfers," *American Economic Review*, 1988, *78* (4), 691–700.

Bolton, Patrick and Mathias Dewatripont, *Contract Theory*, Cambridge, MA: MIT Press, 2005.

Bonell, Michael Joachim, *An International Restatement Of Contract Law: The UNIDROIT Principles Of International Commercial Contracts*, Irvington-on-Hudson, NY: Transnational Publishers, 1997.

Boolos, George S., John P. Burgess, and Richard C. Jeffrey, *Computability and Logic*, 4th ed., Cambridge, England: Cambridge University Press, 2002.

Bork, Robert H., *The Antitrust Paradox*, New York: Basic Books, 1978.

Buchanan, James M., "In Defense of Caveat Emptor," *University of Chicago Law Review*, 1970, *38* (1), 64–73.

Bull, Clive, "The Existence of Self-Enforcing Implicit Contracts," *Quarterly Journal of Economics*, 1987, *102* (1), 147–160.

Caillaud, Bernard and Benjamin E. Hermalin, "Hidden Action and Incentives," Technical Report, University of California, Berkeley, http://faculty.haas.berkeley.edu/hermalin/agencyread.pdf 2000.

_ **and** _ , "Hidden-Information Agency," Technical Report, University of California, Berkeley, http://faculty.haas.berkeley.edu/hermalin/mechread.pdf 2000.

Calabresi, Guido and A. Douglas Melamed, "Property Rules, Liability Rules, and Inalienability: One View of the Cathedral," *Harvard Law Review*, 1972, *85* (6), 1089–1128.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew Rabin, "Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism," *University of Pennsylvania Law Review*, 2003, *151* (3), 1211–1254.

Charny, David, "Nonlegal Sanctions in Commercial Relationships," *Harvard Law Review*, December 1990, *104* (2), 373–467.

_ , "Hypothetical Bargains: The Normative Structure of Contract Interpretation," *Michigan Law Review*, 1991, *89* (7), 1815–1879.

Che, Yeon-Koo and Donald B. Hausch, "Cooperative Investments and the Value of Contracting," *American Economic Review*, 1999, *89* (1), 125–147.

Cheung, Steven N.S., "Transaction Costs, Risk Aversion, and the Choice of Contractual Arrangement," *Journal of Law & Economics*, 1969, *12* (1), 23–42.

Chirelstein, Marvin A., *Concepts and Case Analysis in the Law of Contracts*, New York: Foundation Press, 2006.

Cho, In-Koo and David M. Kreps, "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, May 1987, *102* (2), 179—222.

Chung, Tai-Yeong, "Incomplete Contracts, Specific Investments, and Risk Sharing," *Review of Economic Studies*, 1991, *58* (5), 1031–1042.

Clarkson, Kenneth W., Roger LeRoy Miller, and Timothy J. Muris, "Liquidated Damages vs. Penalties: Sense or Nonsense?," *Wisconsin Law Review*, 1978, (2), 351–390.

Coase, Ronald H., "The Problem of Social Cost," *Journal of Law and Economics*, April 1960, *3* (1), 1–44.

Cook, Philip J. and Daniel A. Graham, "The Demand for Insurance and Protection: The Case of Irreplaceable Commodities," *Quarterly Journal of Economics*, February 1977, *91* (1), 143–156.

Cooter, Robert D., "The Cost of Coase," *Journal of Legal Studies*, 1982, *11* (1), 1–31.

_ , "Unity in Tort, Contract, and Property: The Model of Precaution," *California Law Review*, January 1985, *73* (1), 1–51.

_ **and Melvin A. Eisenberg**, "Damages for Breach of Contract," *California Law Review*, October 1985, *73* (5), 1432–1481.

_ **, Stephen Marks, and Robert Mnookin**, "Bargaining in the Shadow of the Law: A Testable Model of Strategic Behavior," *Journal of Legal Studies*, 1982, *11* (2), 225–251.

Craswell, Richard, "Precontractual Investigation as an Optimal Precaution Problem," *Journal of Legal Studies*, June 1988, *17* (2), 401–436.

_ , "Contract Law, Default Rules, and the Philosophy of Promising," *Michigan Law Review*, 1989, *88* (3), 489–529.

_ , "Performance, Reliance, and One-Sided Information," *Journal of Legal Studies*, June 1989, *18* (2), 365–401.

_ , "Insecurity, Repudiation, and Cure," *Journal of Legal Studies*, June 1990, *19* (2), 399–434.

_ , "Passing on the Costs of Legal Rules: Efficiency and Distribution in Buyer-Seller Relationships," *Stanford Law Review*, 1991, *43* (2), 361–398.

_ , "Property Rules and Liability Rules in Unconscionability and Related Doctrines," *University of Chicago Law Review*, 1993, *60* (1), 1–65.

_ , "Damage Multipliers in Market Relationships," *Journal of Legal Studies*, June 1996, *25* (2), 463–492.

_ , "Offer, Acceptance, and Efficient Reliance," *Stanford Law Review*, 1996, *48* (3), 481–553.

_ , "Deterrence and Damages: The Multiplier Principle and Its Alternatives," *Michigan Law Review*, June 1999, *97* (7), 2185–2238.

_ , "Do Trade Customs Exist?," in Jody S. Kraus and Steven D. Walt, eds., *The Jurisprudential Foundations of Corporate and Commercial Law*, Cambridge: Cambridge University Press, 2000.

_ , "In that Case, What is the Question? Economics and the Demands of Contract Theory," *Yale Law Journal*, January 2003, *112* (4), 903–924.

_ , "Misrepresentation and Nondisclosure in Contract Law and Elsewhere," *Virginia Law Review*, in press.

Crocker, Keith and Scott E. Masten, "Mitigating Contractual Hazards: Unilateral Options and Contract Length," *RAND Journal of Economics*, 1988, *19* (3), 327–343.

Cungu, Azeta and Johan Swinnen, "Investment and Contract Enforcement in Transition: Evidence from Hungary," Technical Report 127, LICOS Centre for Transition Economics, Katholieke Universiteit Leuven, Leuven, Belgium 2003.

Dal Bó, Ernesto, "Bribing Voters," Working Paper, University of California, Berkeley 2003.

Darby, Michael R. and Edi Karni, "Free Competition and the Optimal Amount of Fraud," *Journal of Law and Economics*, April 1973, *16* (1), 67–88.

Debreu, Gerard, *Theory of Value*, New Haven, CT: Yale University Press, 1959.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini, "Recent Developments in Modeling Unforeseen Contingencies," *European Economic Review*, 1998, *42* (3), 523–542.

_ , _ , and _ , "Representing Preferences with a Unique Subjective State Space," *Econometrica*, July 2001, *69* (4), 891–934.

DellaVigna, Stefano and Ulrike Malmendier, "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, May 2004, *119* (2), 353–402.

_ and _ , "Paying not to go to the Gym," *American Economic Review*, in press.

Demski, Joel S. and David E.M. Sappington, "Resolving Double Moral Hazard Problems with Buyout Agreements," *RAND Journal of Economics*, 1991, *22* (2), 232–240.

Diamond, Peter A., "Search Theory," in J. Eatwell, M. Milgate, and P. Newman, eds., *The New Palgrave: A Dictionary of Economics*, New York: Macmillan, 1987.

_ and Eric Maskin, "An Equilibrium Analysis of Search and Breach of Contract I: Steady States," *Bell Journal of Economics*, Spring 1979, *10* (1), 282–316.

_ and _ , "An Equilibrium Analysis of Search and Breach of Contract II: A Non-Steady State Example," *Journal of Economic Theory*, October 1981, *25* (2), 165–195.

Dworkin, Ronald M., "Is Wealth a Value?," *Journal of Legal Studies*, 1980, *9* (2), 191–226.

Dye, Ronald A., "Costly Contract Contingencies," *International Economic Review*, February 1985, *26* (1), 233–250.

Easterbrook, Frank H., "The Supreme Court, 1983 Term Foreword: The Court and the Economic System," *Harvard Law Review*, 1984, *98* (1), 4–60.

Edlin, Aaron S. and Alan Schwartz, "Optimal Penalties in Contracts," *Chicago-Kent Law Review*, 2003, *78* (1), 33–54.

_ and Benjamin E. Hermalin, "Contract Renegotiation and Options in Agency Problems," *Journal of Law, Economics, and Organization*, 2000, *16* (2), 395–423.

_ and _ , "Implementing the First Best in an Agency Relationship with Renegotiation: A Corrigendum," *Econometrica*, September 2001, *69* (5), 1391–1395.

_ and Stefan Reichelstein, "Holdups, Standard Breach Remedies, and Optimal Investments," *American Economic Review*, 1996, *86* (3), 478–501.

Eisenberg, Melvin A., "The Limits of Cognition and the Limits of Contracts," *Stanford Law Review*, January 1995, *47* (2), 211–259.

Epstein, Richard, "Unconscionability: A Critical Appraisal," *Journal of Law & Economics*, 1975, *18*, 293–315.

Farnsworth, E. Allan, "Your Loss or my Gain? The Dilemma of the Disgorgement Principle in Breach of Contract," *Yale Law Journal*, 1985, *94* (6), 1339–1393.

_ , *Farnsworth on Contracts*, 3rd ed., New York: Aspen Publishers, 2004.

Farrell, Joseph, "Cheap Talk, Coordination, and Entry," *RAND Journal of Economics*, 1987, *18* (1), 34–39.

_ , "Information and the Coase Theorem," *Journal of Economic Perspectives*, Fall 1987, *1* (2), 113–129.

_ , "Meaning and Credibility in Cheap Talk Games," *Games and Economic Behavior*, 1993, *5* (4), 514–531.

Fishman, Michael J. and Kathleen M. Hagerty, "Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers," *Journal of Law, Economics, and Organization*, 2003, *19* (1), 45–63.

Fong, Yuk-fai, "When do Experts Cheat and Whom do They Target?," *RAND Journal of Economics*, Spring 2005, *36* (1), 113–130.

Fried, Charles, *Contract as Promise: A Theory of Contractual Obligation*, Cambridge, MA: Harvard University Press, 1981.

Friedmann, Daniel, "The Efficient Breach Fallacy," *Journal of Legal Studies*, 1989, *18* (1), 1–24.

Fudenberg, Drew and Jean Tirole, "Moral Hazard and Renegotiation in Agency Contracts," *Econometrica*, November 1990, *58* (6), 1279–1319.

_ and _ , *Game Theory*, Cambridge, MA: MIT Press, 1991.

Fuller, Lon L., "Consideration as Form," *Columbia Law Review*, 1941, *41* (5), 799–824.

_ and William R. Perdue, "The Reliance Interest in Contract Damages," *Yale Law Journal*, November 1936, *46* (1), 52–96. January 1937, *46* (3), 373–420.

Gergen, Mark, "The Use of Open Terms in Contract," *Columbia Law Review*, 1992, *92* (5), 997–1081.

Gibbons, Robert, *Game Theory for Applied Economists*, Princeton, NJ: Princeton University Press, 1992.

Goetz, Charles J. and Robert E. Scott, "Liquidated Damages, Penalties and the Just Compensation Principle: Some Notes on an Enforcement Model and a Theory of Efficient Breach," *Columbia Law Review*, May 1977, *77* (4), 554–594.

_ and _, "Enforcing Promises: An Examination of the Basis of Contract," *Yale Law Journal*, 1980, *89* (7), 1261–1322.

_ and _, "Principles of Relational Contracts," *Virginia Law Review*, 1981, *67* (6), 1089–1150.

_ and _, "The Mitigation Principle: Toward a General Theory of Contractual Obligation," *Virginia Law Review*, September 1983, *69* (6), 967–1024.

_ and _, "The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms," *California Law Review*, 1985, *73* (2), 261–322.

Goldberg, Victor P., "Bloomer Girl Revisited or How to Frame an Unmade Picture," *Wisconsin Law Review*, 1998, *1998* (4), 1051–1084.

_, "In Search of Best Efforts: Reinterpreting Bloor v. Falstaff," *St. Louis University Law Journal*, 2000, *44* (4), 1465–1485.

_ and John R. Erickson, "Quantity and Price Adjustment in Long-term Contracts: A Case Study of Petroleum Coke," *Journal of Law & Economics*, 1987, *30* (2), 369–398.

Greif, Avner and Eugene Kandel, "Contract Enforcement Institutions: Historical Perspective and Current Status in Russia," in Edward P. Lazear, ed., *Economic Transition in Eastern Europe and Russia: Realities of Reform*, Stanford, CA: Hoover Institution Press, 1995, pp. 291–321.

Grossman, Sanford J., "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, 1981, *21*, 461–83.

_ and Oliver D. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica*, January 1983, *51* (1), 7–46.

_ and _, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, August 1986, *94* (4), 691–719.

Güth, Werner, Rolf Schmittberger, and Bernd Schwarze, "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, 1982, *3* (4), 367–388.

Hadfield, Gillian K., "Problematic Relations: Franchising and the Law of Incomplete Contracts," *Stanford Law Review*, 1990, *42* (4), 927–992.

_, "Judicial Competence and the Interpretation of Incomplete Contracts," *Journal of Legal Studies*, 1994, *23* (1), 159–184.

Harris, Milton and Artur Raviv, "Some Results on Incentive Contracts with Applications to Education and Employment, Health Insurance, and Law Enforcement," *American Economic Review*, March 1978, *68* (1), 20–30.

Hart, Oliver D., "Incomplete Contracts," in J. Eatwell, M. Milgate, and P. Newman, eds., *Allocation, Information, and Markets*, New York: W.W. Norton & Co., 1987.

_, "An Economist's Perspective on the Theory of the Firm," *Columbia Law Review*, 1989, *89* (7), 1757–1774.

_ and John H. Moore, "Default and Renegotiation: A Dynamic Model of Debt," *Quarterly Journal of Economics*, February 1998, *113* (1), 1–41.

_ and _, "Foundations of Incomplete Contracts," *Review of Economic Studies*, January 1999, *66* (1), 115–138.

_ and _, "Agreeing Now to Agree Later: Contracts that Rule Out but do not Rule In," 2004. Harvard Law and Economics Discussion Paper No. 465.

Hermalin, Benjamin E., "An Economic Analysis of Takings," *Journal of Law, Economics, and Organization*, 1995, *11* (1), 64–86.

_, "Economics and Corporate Culture," in Cary L. Cooper, Sue Cartwright, and P. Christopher Earley, eds., *The International Handbook of Organizational Culture and Climate*, Chichester, England: John Wiley & Sons, Ltd., 2001.

_, "Adverse Selection, Short-Term Contracting, and the Underprovision of On-the-Job Training," *Contributions to Economic Analysis & Policy*, 2002, *1* (1), Article 5.

_ and Michael L. Katz, "Moral Hazard and Verifiability: The Effects of Renegotiation in Agency," *Econometrica*, November 1991, *59* (6), 1735–1753.

_ and _, "Judicial Modification of Contracts Between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach," *Journal of Law, Economics, and Organization*, Fall 1993, *9* (2), 230–255.

_ and _, "Privacy, Property Rights, & Efficiency: The Economics of Privacy as Secrecy," *Quantitative Marketing and Economics*, in press.

Hillman, Robert A., "Policing Contract Modifications Under the UCC: Good Faith and the Doctrine of Economic Duress," *Iowa Law Review*, 1979, *64* (4), 849–902.

Holmes, Oliver Wendell, "The Path of the Law," *Harvard Law Review*, 1897, *10* (8), 457–478.

Holmstrom, Bengt, "Moral Hazard in Teams," *Bell Journal of Economics*, Autumn 1982, *13* (2), 324–340.

Honnold, John O., *Uniform Law For International Sales Under The 1980 United Nations Convention*, Cambridge, MA: Kluwer Law International, 1999.

Jackson, Thomas H., "Anticipatory Repudiation and the Temporal Element of Contract Law: An Economic Inquiry into Contract Damages in Cases of Prospective Nonperformance," *Stanford Law Review*, November 1978, *31* (1), 69–119.

Johnston, Jason Scott, "Strategic Bargaining and the Economic Theory of Contract Default Rules," *Yale Law Journal*, December 1990, *100* (3), 615–664.

Jolls, Christine M., "Contracts as Bilateral Commitments: A New Perspective on Contract Modification," *Journal of Legal Studies*, 1997, *26* (1), 203–237.

_ , Cass R. Sunstein, and Richard Thaler, "A Behavioral Approach to Law and Economics," *Stanford Law Review*, May 1998, *50* (5), 1471–1550.

Joskow, Paul L., "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets," *American Economic Review*, 1987, *77* (1), 168–85.

Kahan, Marcel and Michael Klausner, "Standardization and Innovation in Corporate Contracting (or The Economics of Boilerplate)," *Virginia Law Review*, 1997, *83* (4), 713–770.

Kaplow, Louis, "Rules Versus Standards: An Economic Analysis," *Duke Law Journal*, 1992, *42* (3), 557–629.

_ , "A Model of the Optimal Complexity of Legal Rules," *Journal of Law, Economics, & Organization*, April 1995, *11* (1), 150–163.

_ and Steven Shavell, "Property Rules Versus Liability Rules: An Economic Analysis," *Harvard Law Review*, February 1996, *109* (4), 713–790.

Katz, Avery W., "The Strategic Structure of Offer and Acceptance: Game Theory and the Law of Contract Formation," *Michigan Law Review*, November 1990, *89* (2), 215–295.

_ , "Your Terms or Mine? The Duty to Read the Fine Print in Contracts," *RAND Journal of Economics*, Winter 1990, *21* (4), 518–537.

_ , "Transaction Costs and the Legal Mechanics of Contract Formation: When Should Silence in the Face of an Offer be Construed as Acceptance?," *Journal of Law, Economics, and Organization*, April 1993, *9* (1), 77–97.

_ , "Positivism and the Separation of Law and Economics," *Michigan Law Review*, 1996, *94* (7), 2229–2269.

_ , "Taking Private Ordering Seriously," *University of Pennsylvania Law Review*, May 1996, *144* (5), 1745–1763.

_ , "When Should an Offer Stick? The Economics of Promissory Estoppel in Preliminary Negotiations," *Yale Law Journal*, 1996, *105* (5), 1249–1309.

_ , "An Economic Analysis of the Guaranty Contract," *University of Chicago Law Review*, 1999, *66* (1), 47–116.

_ , "The Economics of Form and Substance in Contract Interpretation," *Columbia Law Review*, March 2004, *103* (2), 496–538.

Kelman, Steven, *What Price Incentives? Economists and the Environment*, Boston: Auburn House Publishing Co., 1981.

Kennedy, Duncan, "Distributive and Paternalist Motives in Contract and Tort Law, with Special Reference to Compulsory Terms and Unequal Bargaining Power," *Maryland Law Review*, 1982, *41* (4), 563–658.

Kenney, Roy W. and Benjamin Klein, "The Economics of Block Booking," *Journal of Law & Economics*, 1983, *26* (3), 497–540.

_ and _ , "How Block Booking Facilitated Self-Enforcing Film Contracts," *Journal of Law & Economics*, 2000, *43* (2), 427–435.

Kessler, Friedrich, "Contracts of Adhesion—Some Thoughts about Freedom of Contract," *Columbia Law Review*, 1943, *43* (5), 629–642.

Klein, Benjamin, "Transaction Cost Determinants of Unfair Contractual Arrangements," *American Economic Review*, May 1980, *70* (2), 356–360.

_ and Keith Leffler, "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy*, 1981, *89* (4), 615–641.

_ and Kevin Murphy, "Vertical Restraints as Contract Enforcement Mechanisms," *Journal of Law & Economics*, 1988, *31* (2), 265–297.

Kornhauser, Lewis A., "Reliance, Reputation, and Breach of Contract," *Journal of Law and Economics*, October 1983, *26* (2), 691–706.

Korobkin, Russell B., "Inertia and Preference in Contract Negotiation: The Psychological Power of Default Rules and Form Terms," *Vanderbilt Law Review*, 1998, *51* (6), 1583–1652.

_ , "The Status Quo Bias and Contract Default Rules," *Cornell Law Review*, 1998, *83* (3), 608–687.

_ **and Thomas S. Ulen**, "Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics," *California Law Review*, July 2000, *88* (4), 1051–1142.

Krasa, Stefan and Anne P. Villamil, "Optimal Contracts When Enforcement is a Decision Variable," *Econometrica*, January 2000, *68* (1), 119–134.

Kraus, Jody S., "Decoupling Sales Law from the Acceptance-Rejection Fulcrum," *Yale Law Journal*, October 1994, *104* (1), 129–205.

Kreps, David M., *A Course in Microeconomic Theory*, Princeton, NJ: Princeton University Press, 1990.

Kronman, Anthony T., "Mistake, Disclosure, Information, and the Law of Contracts," *Journal of Legal Studies*, 1978, *7* (1), 1–34.

_ , "Specific Performance," *University of Chicago Law Review*, Winter 1978, *45* (2), 351–382.

_ , "Living in the Law," *University of Chicago Law Review*, 1987, *54* (3), 835–876.

Kull, Andrew, "Mistake, Frustration, and the Windfall Principle of Contract Remedies," *Hastings Law Journal*, November 1991, *43* (1), 1–55.

Laffont, Jean-Jacques, *Fundamentals of Public Economics*, Cambridge, MA: MIT Press, 1988.

_ **and David Martimont**, *The Theory of Incentives*, Princeton, NJ: Princeton University Press, 2002.

_ **and Jean Tirole**, "The Dynamics of Incentive Contracts," *Econometrica*, 1988, *56* (5), 1153–1175.

_ **and** _ , *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press, 1993.

Landa, Janet T., "A Theory of the Ethnically Homogeneous Middleman Group: An Institutional Alternative to Contract Law," *Journal of Legal Studies*, June 1981, *10* (2), 349–362.

Lando, Henrik and Caspar Rose, "On The Enforcement Of Specific Performance In Civil Law Countries," *International Review of Law and Economics*, 2004, *24* (4), 473–487.

Lando, Ole and Hugh Beale, eds, *Contract Law*, The Hague: Kluwer Law International, 2002.

Levin, Jonathan, "Relational Incentive Contracts," *American Economic Review*, June 2003, *93* (3), 835–857.

Llewellyn, Karl N., *The Common Law Tradition: Deciding Appeals*, Boston: Little, Brown and Co., 1960.

MacLeod, W. Bentley and James M. Malcolmson, "Investments, Holdup, and the Form of Market Contracts," *American Economic Review*, 1993, *83* (4), 811–837.

Macneil, Ian R., "Efficient Breach of Contract: Circles in the Sky," *Virginia Law Review*, May 1982, *68* (5), 947–969.

Mankiw, N. Gregory and Michael D. Whinston, "Free Entry and Social Inefficiency," *RAND Journal of Economics*, 1986, *17* (1), 48–58.

Mann, Ronald J., "Explaining the Pattern of Secured Credit," *Harvard Law Review*, 1997, *110* (3), 625–683.

_ , "The Role of Secured Credit in Small-Business Lending," *Georgetown Law Journal*, 1997, *86* (1), 1–44.

_ , "Verification Institutions in Financing Transactions," *Georgetown Law Journal*, 1999, *87* (7), 2225–2272.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green, *Microeconomic Theory*, Oxford: Oxford University Press, 1995.

Maskin, Eric and Jean Tirole, "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, January 1999, *66* (1), 83–114.

Masten, Scott E. and Edward A. Snyder, "United States versus United Shoe Machinery Corporation: On the Merits," *Journal of Law & Economics*, 1993, *36* (1), 33–70.

Matthews, Steven and Andrew Postlethwaite, "Quality Testing and Disclosure," *RAND Journal of Economics*, 1985, *16* (3), 328–340.

_ **and John H. Moore**, "Monopoly Provision of Quality and Warranties: An Exploration in the Theory of Multidimensional Screening," *Econometrica*, March 1987, *55* (2), 441–467.

Medema, Steven G. and Richard O. Zerbe, "The Coase Theorem," in Boudewijn Bouckaert and Gerrit De Geest, eds., *Encyclopedia of Law and Economics*, Vol. 1, Cheltenham, England: Edward Elgar, 2000, chapter 0730, pp. 836–892.

Menell, Peter S., "A Note on Private versus Social Incentives to Sue in a Costly Legal System," *Journal of Legal Studies*, 1983, *12* (1), 41–52.

Muris, Timothy J., "The Costs of Freely Granting Specific Performance," *Duke Law Journal*, December 1982, (6), 1053–1069.

_ , "Cost of Completion or Diminution in Market Value: The Relevance of Subjective Value," *Journal of Legal Studies*, June 1983, *12* (2), 379–400.

Myerson, Roger B., "Justice, Institutions, and Multiple Equilibria," *Chicago Journal of International Law*, 2004, *5* (1), 91–108.

Nash, John F., "The Bargaining Problem," *Econometrica*, April 1950, *18* (2), 155–162.

_ , "Noncooperative Games," *Annals of Mathematics*, 1951, *54* (2), 286–295.

Nöldeke, Georg and Klaus M. Schmidt, "Option Contracts and Renegotiation: A Solution to the Hold-up Problem," *RAND Journal of Economics*, 1995, *26* (2), 163–179.

_ **and** _ , "Sequential Investments and Options to Own," *RAND Journal of Economics*, 1998, *29* (4), 633–653.

Okun, Arthur M., *Equality and Efficiency: The Big Tradeoff*, Washington, DC: The Brookings Institution, 1975.

Pildes, Richard and Elizabeth Anderson, "Slinging Arrows at Democracy: Social Choice Theory, Value Pluralism, and Democratic Politics," *Columbia Law Review*, 1990, *90* (8), 2121–2214.

Pirrong, Stephen Craig, "Contracting Practices in Bulk Shipping Markets: A Transactions Cost Explanation," *Journal of Law & Economics*, 1993, *36* (2), 937–976.

Polinsky, A. Mitchell, "Economic Analysis as a Potentially Defective Product: A Buyer's Guide to Posner's Economic Analysis of Law," *Harvard Law Review*, 1974, *87* (8), 1655–1681.

_ , "Risk Sharing Through Breach of Contract Remedies," *Journal of Legal Studies*, June 1983, *12* (2), 427–444.

_ **and Steven Shavell**, "Punitive Damages: An Economic Analysis," *Harvard Law Review*, February 1998, *111* (4), 869–962.

Posner, Eric A., "Contract Law in the Welfare State: A Defense of the Unconscionability Doctrine, Usury Laws, and Related Limitations on the Freedom to Contract," *Journal of Legal Studies*, 1995, *24* (2), 283–319.

_ , "The Parol Evidence Rule, the Plain Meaning Rule, and the Principles of Contractual Interpretation," *University of Pennsylvania Law Review*, 1998, *146* (2), 533–577.

_ , "Economic Analysis of Contract Law After Three Decades: Success or Failure?," *Yale Law Journal*, January 2003, *112* (4), 829–880.

_ , "There Are No Penalty Default Rules in Contract Law," 2005. University of Chicago Law & Economics, Olin Working Paper No. 237.

Posner, Richard A., "Taxation by Regulation," *Bell Journal of Economics*, 1971, *2* (1), 22–50.

_ , *Antitrust Law: An Economic Perspective*, Chicago: University of Chicago Press, 1976.

_ , "Gratuitous Promises in Economics and Law," *Journal of Legal Studies*, 1977, *6* (2), 411–426.

_ , *Economic Analysis of Law*, 6th ed., Boston: Little, Brown & Co., 2003.

_ **and Andrew M. Rosenfield**, "Impossibility and Related Doctrines in Contract Law: An Economic Analysis," *Journal of Legal Studies*, 1977, *6* (1), 83–118.

Priest, George L., "Breach and Remedy for the Tender of Nonconforming Goods Under the Uniform Commercial Code: An Economic Approach," *Harvard Law Review*, 1978, *91* (5), 960–1001.

Quillen, Gwyn D., "Contract Damages and Cross-Subsidization," *Southern California Law Review*, May 1988, *61* (4), 1125–1141.

Rabin, Matthew, "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, December 1993, *83* (5), 1281–1302.

_ , "Psychology and Economics," *Journal of Economic Literature*, March 1998, *36* (1), 11–46.

Rachlinski, Jeffrey J., "A Positive Psychological Theory of Judging in Hindsight," *University of Chicago Law Review*, 1998, *65* (2), 571–625.

Radin, Margaret Jane, "Market-Inalienability," *Harvard Law Review*, 1987, *100* (8), 1849–1937.

Rakoff, Todd, "Contracts of Adhesion: An Essay in Reconstruction," *Harvard Law Review*, 1983, *96* (6), 1173–1284.

Rasmusen, Eric B., "Explaining Incomplete Contracts as the Result of Contract-Reading Costs," *Advances in Economic Analysis & Policy*, 2001, *1* (1), Article 2.

_ **, J. Mark Ramseyer, and John S. Wiley**, "Naked Exclusion," *American Economic Review*, December 1991, *81* (5), 1137–1145.

Rea, Samuel A., "Nonpecuniary Loss and Breach of Contract," *Journal of Legal Studies*, January 1982, *11* (1), 35–53.

_ , "Efficiency Implications of Penalties and Liquidated Damages," *Journal of Legal Studies*, January 1984, *13* (1), 147–167.

Rogerson, William P., "Efficient Reliance and Damage Measures for Breach of Contract," *RAND Journal of Economics*, 1984, *15* (1), 39–53.

_ , "Contractual Solutions to the Hold-up Problem," *Review of Economic Studies*, 1992, *59* (4), 777–793.

Rubinstein, Ariel, *Modeling Bounded Rationality* Zeuthen Lecture Book series, Cambridge, MA: MIT Press, 1998.

Samuelson, William, "A Comment on the Coase Theorem," in Alvin E. Roth, ed., *Game-Theoretic Models of Bargaining*, Cambridge: Cambridge University Press, 1985, pp. 321–339.

Schelling, Thomas C., *The Strategy of Conflict*, Cambridge, MA: Harvard University Press, 1960.

Schlechtreim, Peter, *Commentary on the UN Convention on the International Sale of Goods*, Oxford: Clarendon Press, 1998.

Schwartz, Alan, "The Case for Specific Performance," *Yale Law Journal*, December 1979, *89* (2), 271–306.

_ , "A Theory of Loan Priorities," *Journal of Legal Studies*, 1989, *18* (2), 209–261.

_ , "The Myth that Promisees Prefer Supracompensatory Remedies: An Analysis of Contracting for Damage Measures," *Yale Law Journal*, November 1990, *100* (2), 369–407.

_ , "Relational Contracts In The Courts: An Analysis Of Incomplete Agreements And Judicial Strategies," *Journal of Legal Studies*, 1992, *21* (2), 271–318.

_ **and Joel Watson**, "The Law and Economics of Costly Contracting," *Journal of Law, Economics, and Organization*, 2003, *20* (1), 2–31.

_ **and Louis L. Wilde**, "Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis," *University of Pennsylvania Law Review*, 1979, *127* (3), 630–682.

_ **and Robert E. Scott**, "Contract Theory and the Limits of Contract Law," *Yale Law Journal*, 2004, *113* (3), 541–620.

Schweizer, Urs, "Externalities and the Coase Theorem: Hypothesis or Result?," *Journal of Institutional and Theoretical Economics*, 1988, *144* (2), 245–266.

Scott, Robert E., "A Relational Theory of Secured Financing," *Columbia Law Review*, 1986, *86* (5), 901–977.

_ , "The Case for Formalism in Relational Contract," *Northwestern University Law Review*, 2000, *94* (3), 847–876.

_ , "The Rise and Fall of Article 2," *Louisiana Law Review*, 2002, *62* (4), 1009–1064.

_ **and George Triantis**, "Anticipating Litigation in Contract Design," *Yale Law Journal*, January 2006, *115* (4), 814–879.

Segal, Ilya R., "Complexity and Renegotiation: A Foundation for Incomplete Contracts," *Review of Economic Studies*, January 1999, *66* (1), 57–82.

_ , "Contracting with Externalities," *Quarterly Journal of Economics*, May 1999, *114* (2), 337–388.

_ **and Michael D. Whinston**, "Naked Exclusion: Comment," *American Economic Review*, March 2000, *90* (1), 296–309.

_ **and** _ , "Robust Predictions for Bilateral Contracting with Externalities," *Econometrica*, May 2003, *71* (3), 757–791.

Sen, Amartya, "The Impossibility of a Paretian Liberal," *Journal of Political Economy*, 1970, *78* (1), 152–157.

Shapiro, Carl, "Premiums for High Quality Products as Returns to Reputation," *Quarterly Journal of Economics*, 1983, *98* (4), 659–680.

Shavell, Steven, "Damage Measures for Breach of Contract," *Bell Journal of Economics*, 1980, *11* (2), 466–490.

_ , "The Design of Contracts and Remedies for Breach," *Quarterly Journal of Economics*, February 1984, *99* (1), 121–148.

_ , "An Economic Analysis of Altruism and Deferred Gifts," *Journal of Legal Studies*, 1991, *20* (2), 401–421.

_ , "Acquisition and Disclosure of Information Prior to Sale," *RAND Journal of Economics*, 1994, *25* (1), 20–36.

_ , "Contracts, Holdup, and Legal Intervention," March 2005. Harvard Law School Discussion Paper.

_ , "Specific Performance versus Damages for Breach of Contract: An Economic Analysis," *Texas Law Review*, March 2006, *84* (4), 831–876.

_ , "On the Writing and the Interpretation of Contracts," *Journal of Law, Economics, and Organization*, in press.

_ **and Louis Kaplow**, *Fairness versus Welfare*, Cambridge, MA: Harvard University Press, 2002.

Shepard, Andrea, "Licensing to Enhance Demand for New Technologies," *RAND Journal of Economics*, 1987, *18* (3), 360–368.

Simon, Herbert A., "Theories of Bounded Rationality," in C.B. McGuire and R. Radner, eds., *Decision and Organization: A volume in Honor of Jacob Marschak*, Amsterdam: North Holland, 1972.

Spence, A. Michael, "Job Market Signaling," *Quarterly Journal of Economics*, 1973, *87* (3), 355–374.

_ , "Monopoly, Quality, and Regulation," *Bell Journal of Economics*, 1975, *6* (2), 417–429.

Spier, Kathryn E., "Incomplete Contracts and Signalling," *RAND Journal of Economics*, Autumn 1992, *23* (3), 432–443.

_ , "Settlement Bargaining and the Design of Damage Awards," *Journal of Law, Economics, and Organization*, April 1994, *10* (1), 84–95.

_ **and Michael D. Whinston**, "On the Efficiency of Privately Stipulated Damages for Breach of Contract: Entry Barriers, Reliance, and Renegotiation," *RAND Journal of Economics*, 1995, *26* (2), 180–202.

Sutton, John, "Non-Cooperative Bargaining Theory: An Introduction," *Review of Economic Studies*, 1986, *53* (5), 709–724.

Thomas, Jonathan and Tim Worrall, "Self-Enforcing Wage Contracts," *Review of Economic Studies*, 1988, *55* (4), 541–553.

Tirole, Jean, *The Theory of Industrial Organization*, Cambridge, MA: MIT Press, 1988.

_ , "Incomplete Contracts: Where Do We Stand?," *Econometrica*, July 1999, *67* (4), 741–781.

Tobin, James, "On Limiting the Domain of Inequality," *Journal of Law & Economics*, 1970, *13* (2), 263–277.

Townsend, Robert M., "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 1979, *20*, 265–293.

Treitel, G.H., ed., *The Law of Contract*, 11th ed., London: Sweet & Maxwell, 2003.

Triantis, Alexander J. and George G. Triantis, "Timing Problems in Contract Breach Decisions," *Journal of Law and Economics*, April 1998, *41* (1), 163–207.

Triantis, George G., "Secured Debt Under Conditions of Imperfect Information," *Journal of Legal Studies*, 1992, *21* (1), 225–258.

Tullock, Gordon, *The Rent-Seeking Society*, Indianapolis: Liberty Fund, 2005. edited by Charles K. Rowley.

Ulen, Thomas S., "The Efficiency of Specific Performance: Toward a Unified Theory of Contract Remedies," *Michigan Law Review*, November 1984, *83* (2), 341–403.

Varian, Hal R., *Microeconomic Analysis*, 3rd ed., New York: W.W. Norton, 1992.

Walras, Leon, *Eléments d'économie politique pure*, Lausanne: L. Corbaz, 1874.

Webb, David C., "Two-Period Financial Contracts with Private Information and Costly State Verification," *Quarterly Journal of Economics*, August 1992, *107* (3), 1113–1123.

Weinstein, Mark, "Profit Sharing Contracts in Hollywood: Evolution and Analysis," *Journal of Legal Studies*, 1998, *27* (1), 67–112.

White, James J. and Robert S. Summers, *Uniform Commercial Code*, 5th ed., St. Paul: West Group, 2000.

White, Michelle J., "Contract Breach and Contract Discharge due to Impossibility: A Unified Theory," *Journal of Legal Studies*, 1988, *17* (2), 353–376.

Williamson, Oliver E., *Markets and Hierarchies, Analysis and Antitrust Implications: A Study in the Economics of Internal Organization*, New York: The Free Press, 1975.

Willig, Robert D., "Consumer Surplus Without Apology," *American Economic Review*, September 1976, *66* (4), 589–597.

Wittgenstein, Ludwig, *Tractatus Logico-Philosophicus*, London: Routledge and Kegan Paul, 1922.

Wittman, Donald, "Optimal Pricing of Sequential Inputs: Last Clear Chance, Mitigation of Damages, and Related Doctrines in the Law," *Journal of Legal Studies*, January 1981, *10* (1), 65–91.

Wolcher, Louis E., "Price Discrimination and Inefficient Risk Allocation Under the Rule of *Hadley v. Baxendale*," *Research in Law and Economics*, 1989, *9*, 9–31.

Essay

Economic Analysis of Contract Law After Three Decades: Success or Failure?

Eric A. Posner†

INTRODUCTION

Modern economic analysis of contract law began about thirty years ago and, many scholars would agree, has become the dominant academic style of contract theory. Traditional doctrinal analysis exerts less influence than it did prior to 1970 and enjoys little prestige. Philosophical work on the nature of promising has captured some attention, but petered out in the 1980s, with little to show for the effort other than arid generalizations about the nature of promising. Academic critiques from the left no longer stir up excitement as they did twenty years ago. Scholarship influenced by cognitive psychology has so far produced few insights. Only economic analysis seems to be on solid footing.

One way to validate a field's claims is to look at its history. Economically oriented scholars writing in the early 1970s had foundational insights, and then over time subsequent writers have criticized and refined them; because these refinements were derived from common premises, there has been a sense of forward movement in the subject, of the building of an increasingly sophisticated consensus. Although critics of economic analysis deride its scientific aspirations, the steady accumulation of insights over time resembles scientific progress. Doctrinal, philosophical, and critical scholarship by contrast has been static. The authors agree or disagree, and about the same things, as much today as they did twenty or thirty years ago.

Yet there are grounds for concern about the economic analysis of contract law. Careful students of its history know that the sense of convergence ended years ago; in the last ten years, theory has become divergent, and impasses have emerged. The simple models that dominated discussion prior to the 1990s do not predict observed contract doctrine. The more complex models that emerged in the 1980s and dominated discussion in the 1990s failed to predict doctrine or relied on variables that could not, as a practical matter, be measured. As a result, the predictions of these models are indeterminate, and the normative recommendations derived from them are implausible.

For these reasons, I will argue that economic analysis has failed to produce an "economic theory" of contract law, and does not seem likely to be able to do so. By this, I mean that the economic approach does not explain the current system of contract law, nor does it provide a solid basis for criticizing and reforming contract law. This is not to say that the economic approach has not produced any wisdom, but that the nature of its accomplishment turns out to be subtle and will become clear only after an extended discussion.

This Essay has two purposes: to document the failures of economic models to explain contract law or to justify reform, and to provide an explanation for these failures. The explanation centers on the difficulty of developing a model of contractual behavior that can be tested and that does not make unreasonable assumptions about the cognitive abilities of contractual parties.

At the outset, a few comments must be made in order to avoid some possible misunderstandings of the argument. First, I will not argue that some other approach to contract law is superior to the economic approach, nor that economic analysis should be abandoned. If a moral must be extracted from the discussion, it is skepticism about how much additional value economics has to offer to understanding contract law today.

Second, I do not make claims about the value of economic analysis for understanding other areas of law. Indeed, my critique rests on empirical and methodological judgments about the contracts literature, judgments that do not necessarily apply to, say, torts or property. Nor do I take a position in

† Professor of Law, University of Chicago. Thanks to Bill Dodge, Mark Gergen, Richard Posner, Alan Schwartz, Steve Smith, David Snyder, Kathy Spier, and workshop participants at the University of Illinois at Chicago Economics Department for helpful comments, to Bryan Dayton and Tana Ryan for research assistance, and to The Sarah Scaife Foundation Fund and The Lynde and Harry Bradley Foundation Fund for financial support. Richard Craswell's comments on an earlier draft also improved the paper; I am also grateful for his and Ian Ayres's published comments.

this Essay on controversies over the welfarist foundations of economic analysis.[1]

Third, I want to avoid making general arguments about what counts as a good theory. One might argue that any methodology that yields surprises or insights about a familiar topic is valuable, and those surprises or insights should be counted as theories. To avoid these philosophical issues, I will focus on the original aspirations of the economic analysis of contract law: to provide an explanation of existing legal rules, and to provide a basis for criticizing or defending those rules.[2]

Finally, I want to avoid debates about what counts as "economic analysis of contract law" by stipulating that it did not exist before 1970. This is, of course, artificial. Many earlier scholars, including Holmes, Llewellyn, Hale, and Fuller, used economic analysis in the sense that from time to time they would assume that contracting parties are rational and then speculate about how different legal rules would affect these parties' incentives.[3] From a modern perspective, however, their insights seem banal, and that is because post-1970 economic analysis is more systematic and careful.[4] The interesting question is whether the post-1970 commitment to methodological individualism and the other premises of the rational actor approach provide the basis for a theory that can be used to explain or criticize contract law.

My plan is as follows. Part I describes various results from the economic analysis of contract law and compares them with the legal doctrine. In virtually every case, models make either false or indeterminate predictions about the doctrines of contract law. Part II discusses the closely related literature on incomplete contracts, a literature that attempts to

1. For a recent defense of the welfarist approach, see Louis Kaplow & Steven Shavell, *Fairness Versus Welfare*, 114 HARV. L. REV. 961 (2001).

2. This was recently described in RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 26-29 (5th ed. 1998). As Ayres and Craswell point out, authors are more careful today about "explaining" legal rules, but there is no doubt that these authors proffer such explanations frequently, even where the normative project is emphasized. *See* ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 222 (3d ed. 2000) ("The courts reduce the costs of negotiating contracts by supplying efficient default terms."); Kaplow & Shavell, *supra* note 1, at 1163 (arguing that notions of fairness have not "led us seriously astray" from welfare economics, because "basic rules of damages do not seem to reflect such principles").

3. *See* BARBARA H. FRIED, THE PROGRESSIVE ASSAULT ON LAISSEZ FAIRE (1998); Avery Katz, *Reflections on Fuller and Perdue's* The Reliance Interest in Contract Damages*: A Positive Economic Framework*, 21 U. MICH. J.L. REFORM 541 (1988); Richard A. Posner, *Introduction* to THE ESSENTIAL HOLMES, at ix (Richard A. Posner ed., 1992); Alan Schwartz, *Karl Llewellyn and the Origins of Contract Theory*, *in* THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW 12 (Jody S. Kraus & Steven D. Walt eds., 2000).

4. For defenses of the earlier work, see FRIED, *supra* note 3; and Herbert Hovenkamp, *The First Great Law & Economics Movement*, 42 STAN. L. REV. 993 (1990). Fried and Hovenkamp like the earlier work because it struggled with foundational issues and made liberal or progressive recommendations. Although the importance of this scholarship in intellectual history cannot be denied, its lack of continued vitality is almost certainly due to its failure to produce a tractable methodology.

predict the content of contracts, as opposed to contract law. The separation of these two bodies of scholarship, now gradually disappearing, is an accident of history, but useful for seeing the general problems with the economic project. Part III speculates about what went wrong with economic analysis and argues that an ambiguity at the heart of the concept of transaction costs is to blame. Part IV looks at trends in contracts scholarship. Part V criticizes alternative approaches to contract theory.

I. THE ECONOMIC ANALYSIS OF CONTRACT LAW

A. *Premises and Basic Results*

The economic analysis of contract law is too familiar to warrant an extended discussion; there are also several excellent surveys.[5] Fundamental assumptions, common to nearly all efforts at economic analysis, are that individuals have preferences over outcomes, that these preferences obey basic consistency conditions, and that individuals satisfy these preferences subject to an exogenous budget constraint. Contracts scholars usually assume that individuals do not have preferences regarding the consumption or well-being of other individuals, nor regarding contract doctrine itself—there is no preference for expectation damages, for example.[6]

The standard approach assumes that the parties enter a contract in order to secure investment in a jointly beneficial project.[7] The project could be as simple as the sale of a good from Seller to Buyer—with one party (or both) enhancing the gains by an investment that reduces the cost of production for Seller or increases the value of the good for Buyer—or as complex as the construction of a skyscraper. If Buyer can increase the value of the good by making investments prior to delivery, Buyer will want a guarantee that Seller will not increase the price after Seller has observed Buyer's reliance. A contract can sometimes prevent Seller from holding up the Buyer in this way, and thus permit Buyer to invest with knowledge that he will enjoy the full return of his investment.

In their contracts, parties include terms describing performance and governing the main contingencies that affect the value of performance.

5. See the relevant entries in ENCYCLOPEDIA OF LAW AND ECONOMICS (Boudewijn Bouckaert & Gerrit De Geest eds., 2000) and THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW (Peter Newman ed., 1998). *See also* COOTER & ULEN, *supra* note 2, chs. 6-7; POSNER, *supra* note 2, ch. 4; Kaplow & Shavell, *supra* note 1, at 1102-64; Lewis A. Kornhauser, *An Introduction to the Economic Analysis of Contract Remedies*, 57 U. COLO. L. REV. 683 (1986).

6. This is not always true; scholarship on donative promises usually assumes that the promisor cares about the well-being of the promisee. *See, e.g.*, Eric A. Posner, *Altruism, Status, and Trust in the Law of Gifts and Gratuitous Promises*, 1997 WIS. L. REV. 567.

7. Another common rationale for entering a contract is risk sharing. *See* A. Mitchell Polinsky, *Risk Sharing Through Breach of Contract Remedies*, 12 J. LEGAL STUD. 427 (1983).

Terms might describe the goods to be delivered, the date of delivery, or the identity of the party that bears the risk of an accident during the shipment. The terms might also release the seller from its obligation if a strike or similar event occurs. A theoretically complete contract would describe all the possible contingencies, but transaction costs—including the cost of negotiating and writing down the terms—and foreseeing low-probability events, render all contracts incomplete. In addition, parties might choose some terms or avoid others for strategic reasons, in order to exploit superior bargaining power or information asymmetries. Thus, contracts are usually quite incomplete. Parties rely on custom, trade usage, and, in the end, the courts to fill out the terms of the contract.

The terms that appear in contracts, then, depend on what the parties are trying to accomplish, shared understandings about the relevant industry, transaction costs, general characteristics of their interaction such as asymmetric information and unequal bargaining power, and the background legal regime. The last factor, the legal regime, is the focus of the economic analysis of contract law. The question is, broadly speaking, what rules of contract law would best serve the interests of the parties. This question is asked in two different ways, depending on whether the scholar takes a descriptive or a normative approach.

Descriptive analysis provides a "prediction" of contract doctrine. Built into this approach is the assumption that judges decide cases (and/or choose doctrine) in a manner that maximizes efficiency.[8] The question why judges would decide cases in this way, or whether it is necessary for them to do so in order to generate efficient law, is bracketed.[9] The author constructs a model in which parties would maximize their utility if they could enter an optimal contract. They cannot enter such a contract in the absence of legal enforcement, so the question becomes what legal rule enables the parties to enter the optimal contract. This hypothetical legal rule is then compared to

8. The models are usually written as though the concept of efficiency being used is Pareto efficiency: The decisionmaker chooses the rule that maximizes the surplus from cooperation, and, although this might involve the Kaldor-Hicks idea of transferring goods from the person who values them less to the person who values them more, all people who use contracts are better off with an efficient system because prices will reflect the risk of ex post transfers. To be sure, in many cases prices will not adjust, and the transition from an inefficient rule to an efficient rule would likely be a Kaldor-Hicks move, but these will usually be minor considerations.

9. A literature that analyzes this assumption is inconclusive. *See, e.g.*, George L. Priest, *The Common Law Process and the Selection of Efficient Rules*, 6 J. LEGAL STUD. 65 (1977); Paul H. Rubin, *Why Is the Common Law Efficient?*, 6 J. LEGAL STUD. 51 (1977). There is no reason why the prediction must be that contract law is efficient; this seems to be an accident of intellectual history. One could imagine a different theory, along the lines of public choice, that holds that contract law reflects the self-interested decisions of judges to implement policy preferences. Indeed, such an approach has been used by political scientists to explain judicial interpretation of statutes and constitutional provisions. *See, e.g.*, JEFFREY A. SEGAL & HAROLD J. SPAETH, THE SUPREME COURT AND THE ATTITUDINAL MODEL (1993).

actual legal rules, and, if they are the same, the descriptive hypothesis is vindicated.

The normative position assumes that contract law should be efficient. As before, the author constructs a model in which parties can increase their welfare through a contract that is legally enforceable. The author first shows the optimal outcome—where, for example, performance occurs only when the buyer's valuation exceeds the seller's cost, and buyer and seller make efficient investments—and then the equilibrium outcomes under alternative legal rules. Typically, the author recommends one rule as efficient, or shows that different rules are efficient under different assumptions, or else criticizes various existing rules because they do not enable the parties to achieve the optimal outcome.

In the following Sections, I will show the ways in which contract doctrine diverges from the predictions of the descriptive hypotheses, and I will show that the normative implications of the models are weak or nonexistent. The reason for discussing normative and descriptive failures at the same time is that the two are closely connected. From a descriptive perspective, the models generate either false or indeterminate predictions. From a normative perspective, the models generate either implausible or indeterminate recommendations. The reason in both cases is that the determinate models omit important variables, but including these variables makes them indeterminate, or, in some cases, unrealistic, because they place too great a burden on courts. The nature and origin of these difficulties will become clearer as we examine the models.

B. *Remedies*

Much contract doctrine comprises background rules that parties can change, albeit within limits. The victim of breach, by default, receives expectation damages, but the parties can vary this outcome ex ante by providing for liquidated damages in the contract. Their ability to contract around the expectation damages rule in this way is circumscribed by the penalty doctrine, which forbids liquidated damages that are unreasonably high.

At an early stage, scholars argued that the default rule should maximize the ex ante value of the contract. Expectation damages were said to have this effect as a result of an attractive property: They give a party the incentive to breach if and only if the cost of performance for the promisor exceeds the value of performance for the promisee. Performance occurs if

and only if it is efficient. For this reason, expectation damages seemed to be the right measure of damages.[10]

This conclusion was premature, however. First, the argument overlooks the ability of the parties to renegotiate prior to performance. If renegotiation costs are low enough, efficient performance will occur regardless of the remedy. If the remedy is less than expectation damages and performance is efficient, the promisee will bribe the promisor to perform. If the remedy is greater than expectation damages and performance is inefficient, the promisor will pay the promisee for a release.

Second, the argument overlooks the effect of the expectation measure on other incentives. Consider the promisee's incentive to rely or invest in anticipation of performance. Under the rule of expectation damages, the promisee's reliance investment is fully compensated. But if the promisee expects to recover the investment regardless of whether or not trade is efficient, the promisee will overinvest—that is, he will invest as though the return were certain rather than stochastic, externalizing the cost on the promisor.[11] A superior measure of damages would give the promisee the amount of damages that would compensate the promisee if he engaged in efficient reliance, not the amount that would compensate the promisee for the loss given whatever level of reliance was taken.[12]

The concept of efficient investment is subtle, and a numerical example might help. Suppose that Buyer and Seller enter a contract under which Seller promises to supply goods that Buyer needs for his factory. Buyer can increase the value of the goods for his use by investing in adjustments to the factory prior to delivery. Let's say that if Buyer invests 0, his valuation of the goods equals 100. If Buyer invests 5, his valuation of the goods equals 120. If Buyer invests 10, his valuation of the goods equals 128. If Buyer will obtain the goods with certainty, then efficiency requires that he invest 10: $128 - 10 > 120 - 5 > 100 - 0$. However, if Buyer will obtain the goods with only a 50% probability, then efficiency requires that he invest 5: $0.5(120) - 5 > 0.5(128) - 10$, and $0.5(120) - 5 > 0.5(100) - 0$. A person who invests money in some outcome will invest more if the outcome is certain than if the outcome is uncertain. Because expectation damages provide a return to the promisee whether or not breach is efficient, the promisee will invest as though the yield of the investment would occur with probability of 1 rather than with the probability (<1) that performance

10. John H. Barton, *The Economic Basis of Damages for Breach of Contract*, 1 J. LEGAL STUD. 277 (1972); Robert L. Birmingham, *Breach of Contract, Damage Measures, and Economic Efficiency*, 24 RUTGERS L. REV. 273 (1970).

11. *See* William P. Rogerson, *Efficient Reliance and Damage Measures for Breach of Contract*, 15 RAND J. ECON. 39 (1984); Steven Shavell, *The Design of Contracts and Remedies for Breach*, 99 Q.J. ECON. 121 (1984).

12. *See* Robert Cooter, *Unity in Tort, Contract, and Property: The Model of Precaution*, 73 CAL. L. REV. 1, 14 (1985).

occurs. The promisee thus invests an amount greater than would be efficient.[13]

Third, the argument neglects the ability of the parties to design remedial provisions for their contract. If expectation damages are optimal, the parties can achieve the effect of this remedy by giving each side the option to perform or pay an amount that is the function of revealed ex post values. If expectation damages are not optimal, then the parties can choose some superior remedy that would, for example, take account of reliance incentives. These considerations suggest that specific performance of the remedial portion of the contract would be efficient, not expectation damages, which in essence convert the obligation to perform into an option to perform or pay an amount determined by a court.

There are numerous other problems with expectation damages. Expectation damages are also undesirable if courts have trouble determining the parties' valuations at the time of breach. The better remedy is specific performance, which a court can award without determining the promisee's valuation.[14]

Expectation damages are also undesirable when information is asymmetric, unless highly specific conditions are met. Consider the *Hadley* rule, according to which a victim of breach obtains compensation for average, rather than actual, loss, unless he has revealed his valuation to the promisor ex ante.[15] Thus, the shipper cannot recover fully compensatory damages from a carrier who has breached the shipment contract if the shipper does not reveal the specially high value of the goods shipped. The *Hadley* rule gives the shipper an incentive to disclose his valuation prior to contracting, so that the carrier will take optimal precautions given the shipment's value.

But it turns out that the argument can be reversed. Imagine an expansive liability rule that gave the victim of breach actual damages (that is, expectation damages). The defense of *Hadley* implicitly assumed that under the expansive liability rule the high-value shipper would not have an incentive to reveal his valuation: If he is to be fully compensated, he has no

13. For another example, see A. MITCHELL POLINSKY, AN INTRODUCTION TO LAW AND ECONOMICS 32-35 (1983).

14. Kronman argues that the common law efficiently reserves specific performance for disputes involving valuation problems such as those involving unique goods. *See* Anthony Kronman, *Specific Performance*, 45 U. CHI. L. REV. 351 (1978). Schwartz points out that information problems about valuation, enforcement, and so forth are always present, and therefore specific performance should be the default rule. *See* Alan Schwartz, *The Case for Specific Performance*, 89 YALE L.J. 271 (1979); *see also* Thomas S. Ulen, *The Efficiency of Specific Performance: Toward a Unified Theory of Contract Remedies*, 83 MICH. L. REV. 341 (1984). The two remedies also have different effects on reliance incentives. *See* Shavell, *supra* note 11. But, the simplest defense of specific performance is that if parties are rational, they will design an optimal contract, and courts should enforce their terms rather than give the parties an option (expectation damages) when they did not bargain for it.

15. Hadley v. Baxendale, 156 Eng. Rep. 145 (Ex. 1854).

reason to reveal his valuation, which would enable the carrier to charge a higher price. But the expansive liability rule does give the low-value shipper the incentive to reveal his valuation. If he does not, he will be charged ex ante for average compensation, but he would prefer to be charged a lower price, even if this means that the carrier will take less care. If the low-value shippers reveal their valuation, then the carrier can infer that any shipper that does not reveal his valuation must have a high valuation. Both the *Hadley* rule and its opposite give parties incentives to disclose private information.

Authors who have pursued this argument point out that one rule could be better than the other, depending on the distribution of valuations, the cost of revealing information, the relative bargaining power of the party with private information and the uninformed party, and related factors. If there are more low-value shippers than high-value shippers, the expansive liability rule requires more bargaining around, and therefore more transaction costs, and thus might be suboptimal.[16] But the relevant variables are too complex and too hard to determine. We do not observe doctrine incorporating them, nor do we have enough empirical data to be able to guess which rule is based on assumptions that are closer to reality.[17]

16. *See* Ian Ayres & Robert Gertner, *Filling Gaps in Incomplete Contracts: An Economic Theory of Default Rules*, 99 YALE L.J. 87 (1989) [hereinafter Ayres & Gertner, *Filling Gaps*]; Ian Ayres & Robert Gertner, *Strategic Contractual Inefficiency and the Optimal Choice of Legal Rules*, 101 YALE L.J. 729 (1992); Lucian Arye Bebchuk & Steven Shavell, *Information and the Scope of Liability for Breach of Contract: The Rule of* Hadley v. Baxendale, 7 J.L. ECON. & ORG. 284 (1991); Charles J. Goetz & Robert E. Scott, *Enforcing Promises: An Examination of the Basis of Contract*, 89 YALE L.J. 1261, 1299-300 (1980); Jason Scott Johnston, *Strategic Bargaining and the Economic Theory of Contract Default Rules*, 100 YALE L.J. 615 (1990); Louis E. Wolcher, *Price Discrimination and Inefficient Risk Allocation Under the Rule of* Hadley v. Baxendale, *in* 12 RESEARCH IN LAW AND ECONOMICS 9 (Richard O. Zerbe ed., 1989).

17. For further epicycles, see Barry E. Adler, *The Questionable Ascent of* Hadley v. Baxendale, 51 STAN. L. REV. 1547 (1999) (showing that the results of earlier models change if the high type is likely, rather than certain, to suffer a large loss the event of breach). Adler overstates his argument as a critique of *Hadley v. Baxendale* when, in fact, he just shows that courts must take into account yet another factor when determining the optimal rule. More to the point is his skepticism about the possibility that lawmakers could take into account the factors that he identifies when formulating doctrine. *Id.* at 1582. As Bebchuk and Shavell observe in their reply, "Adler does not note any reasons for assuming that the consideration that he discusses involves less practical problems for lawmakers than the considerations on which our analysis has focused." Lucian Arye Bebchuk & Steven Shavell, *Reconsidering Contractual Liability and the Incentive To Reveal Information*, 51 STAN. L. REV. 1615, 1627 (1999). Adler does seem to realize that the accumulating complexities of the analysis undermine its practical value for lawmakers. Bebchuk and Shavell, by contrast, state:

> [O]ur analysis of *Hadley* enables one to recommend that rule with greater confidence than researchers are often able to endorse other legal rules in other contexts. As we explained in some detail, it seems that the *Hadley* rule is clearly desirable for cases (such as *Hadley* itself) in which a minority of buyers has valuations of performance that are substantially higher than the valuations of ordinary buyers.

Id. at 1625. But they do not provide a reason for believing that any of the relevant factors are measurable in general conditions, and one cannot evaluate their historical claim without further evidence.

There are other considerations as well. The remedy that is chosen will affect the incentive of each party to search for the optimal partners prior to contracting, to reveal private information about the probability that performance will be possible, to take precautions against breach, and to renegotiate after information is revealed about the state of the world.[18] Remedies will also affect the ability of the parties to shift risk in a contract when one or both parties are risk-averse. And, as I discuss below, remedies affect the ability of the contracting parties to take advantage of third parties who come onto the scene after the parties have entered the contract and value performance more than either of the contracting parties.

Articles that discuss these various incentives typically bracket most of them for the purpose of analysis and focus on one or two. As a result, the optimal remedy derived from a model is optimal only under narrow conditions. If we are to put the models together and try to draw from them as a group their prediction about contract law, we could take two approaches.

First, we could argue that the models collectively show that different remedies are optimal under different conditions and therefore predict that contract law should incorporate these conditions in doctrine. For example, contract law will make expectation damages the remedy when the parties can make choices only about breach or performance and not about how much to invest. But there are two problems with this approach. The first of these problems is that contract law does not resemble the predictions of the models. Awarding expectation damages is the general rule in contract law, but this rule can be justified by the models only under narrow conditions. Furthermore, doctrine does not make the application of expectation damages turn on variables identified by the models, such as the degree of reliance by the promisee. The second of these problems is that the models taken together are probably indeterminate. To generate predictions, one would need a vast amount of information about the characteristics of the parties and the transactions. If one remedy is best when renegotiation costs are high, and another is best when renegotiation costs are low, we need some way to measure renegotiation costs. If the optimal remedy depends on the shape of probability distributions for sellers' costs and buyers' valuations, we need this information as well. Yet no one has attempted to collect this information, and it is difficult to imagine how this task could be accomplished.

Under the second approach, we could argue that the models collectively show that one particular remedial structure—the existing doctrine of contract law—is optimal given the "average" circumstances of the parties.

18. For a clear discussion, see Richard Craswell, *Contract Remedies, Renegotiation, and the Theory of Efficient Breach*, 61 S. CAL. L. REV. 629 (1988).

We might think, for example, that on average pre-performance investment is not a significant issue, or, if it is, it is adequately controlled by the doctrine of mitigation.[19] The rule of expectation damages is optimal because the perform-or-breach decision matters most, with specific performance reserved for cases where valuation problems are insurmountable. But this view is unsupported by any evidence.

C. *Contract Interpretation*

Many contract disputes turn on questions of interpretation. Seller delivers the goods, but Buyer argues that the goods do not conform to the requirements of the contract. Suppose the contract says "chicken," and the delivery is a scrawny, stewing chicken. Buyer says that "chicken" refers to a plump, juicy broiler; Seller says that the word just identifies the species and leaves the quality of the bird to Seller's discretion.[20] How should the court resolve this dispute?

Economists have proposed a number of interpretive strategies for courts.[21] One is to choose a "majoritarian default," the meaning that most parties to chicken contracts would use, which will often be the same as the customary meaning or trade usage. If parties expect that courts will apply a majoritarian default when disputes arise over the meaning of the contract, they will know that most of the time the court will choose the term that maximizes the probability of efficient trade. Accordingly, they would be more willing to enter a contract in the first place, despite high transaction costs, than they would under an alternative rule. Choosing a majoritarian default rule reduces the negative consequences of high transaction costs.

Another strategy is to choose a "penalty default," a meaning that most parties to chicken contracts would *not* use.[22] This strategy, which would give parties an incentive to write a less ambiguous contract than they might otherwise, has two motivations. First, it discourages parties from externalizing the cost of interpreting the contract on the courts. If parties were clearer, courts would have less work to do. Second, it discourages parties from opportunistically concealing information from each other. If one party knows about the ambiguity of the word "chicken" and prefers the majoritarian meaning, and the other party does not know about the ambiguity, then the first party would have no incentive to disclose the

19. *See* Charles J. Goetz & Robert E. Scott, *The Mitigation Principle: Toward a General Theory of Contractual Obligation*, 69 VA. L. REV. 967 (1983).

20. *Cf.* Frigaliment Importing Co. v. B.N.S. Int'l Sales Corp., 190 F. Supp. 116 (S.D.N.Y. 1960) (holding that the broad meaning of "chicken" is correct).

21. *See, e.g.*, Charles J. Goetz & Robert E. Scott, *The Limits of Expanded Choice: An Analysis of the Interactions Between Express and Implied Contract Terms*, 73 CAL. L. REV. 261, 321 (1985).

22. Ayres & Gertner, *Fillings Gaps*, *supra* note 16, at 95.

ambiguity to the second, unless a penalty default rule held the informed party to the less favorable meaning.

A third strategy is to enforce the contract in a literalistic way. If the party says "chicken" and the dictionary or common sense definition of "chicken" has a general meaning, then Seller has the right to deliver the stewing chicken. The court does not try to determine what most parties mean by "chicken," or what most parties do *not* mean. This strategy, like the penalty default strategy, gives the parties an incentive to be clear, or at least to anticipate how courts normally interpret terms.

A final strategy is for the court to enforce whatever term would be efficient in the particular case. One can derive this term by asking the question: Supposing that transaction costs had been zero at the time of contracting, what would the parties have done? Buyer and Seller would have anticipated their dispute about the meaning of "chicken" and either chosen a more precise term (if trade is still efficient) or not made a deal (if trade is not efficient). What they would have done depends on the costs and values of the various birds. The difference between this strategy and the majoritarian default is the difference between a standard and a rule. The court chooses whatever is efficient for the contract in dispute, rather than enforcing whatever term is efficient for the majority of parties who enter similar or identical contracts.

We have already examined a model comparing the first and second strategy, namely, Ayres and Gertner's model of the *Hadley* rule.[23] The *Hadley* rule, in Ayres and Gertner's argument, plays the role of a penalty default, for they assume that a majority of buyers prefer unlimited liability, which would thus serve as a majoritarian default. Choosing between limited liability and unlimited liability when the contract does not specify one or the other is like choosing between the ordinary meaning of chicken and a narrow meaning of chicken when the contract does not define the term. The choice between these two meanings depends on the same factors that determine the efficiency of the *Hadley* rule: the cost of bargaining around the default rule, the distribution of valuations in the population of buyers, the market power of the seller, the degree to which the seller's performance would improve with superior information, and other factors that are not likely within the grasp of a decisionmaker. Thus, the indeterminacy that afflicts the *Hadley* analysis undermines any effort to choose between a majoritarian and penalty default.

For this reason, one might argue that courts should engage in literalistic enforcement. Indeed, Schwartz makes just such an argument, claiming that the responsibility for choosing default rules puts an unrealistically high

23. *Id.* at 101.

informational burden on the courts.[24] Literalism, by contrast, allows parties to direct courts to enforce obligations that arise under conditions that the courts can verify. But although it is true that literalism does put a lighter burden on courts, it does not follow that literalism is superior to the majoritarian (or penalty) approach. The choice between the two approaches is, as Schwartz acknowledges, an empirical question about which we have no evidence.[25] The most significant problem with Schwartz's analysis, however, is that it depends on the methodological assumption that cognitive limitations do not exist or are minimal. The majoritarian approach depends on the assumption that parties fail to anticipate the future; Schwartz simply assumes the opposite.

This point can also be made about Schwartz's criticism of the view that courts should choose the term that is most efficient in the particular case. Schwartz argues that if the evidence necessary to choose such terms ex post is verifiable, then parties will bargain to the efficient result, in which case judicial intervention is not necessary.[26] In our example, the parties will trade the chicken only if the buyer values it more than the seller does, so that if the buyer accepts the stewing chicken, the ex post interpretation of the contract is effectively that the general meaning of chicken holds. If the evidence is not verifiable, and indeed not observable either, they might bargain to an impasse, or to an inefficient term, in which case courts cannot help. However, if the parties are boundedly rational—again, outside Schwartz's model—we do not know how they would bargain with each other, and therefore whether a court could improve on the outcome.

Let me summarize. From a descriptive perspective, we can distinguish two bodies of work. The standard economic analysis of default rules is broadly consistent with judicial practices; courts employ a mix of majoritarian and penalty defaults. But it does no more than rationalize these practices, for there is no way to measure the variables that determine the relative efficiency of the rules. Schwartz's argument, which is simpler and truer to economic premises, fails to account for courts' refusal (for the most part) to rely on the literalistic approach.[27]

From a normative perspective, Schwartz's argument that courts should engage in literalistic interpretation should appeal to those steeped in

24. Alan Schwartz, *The Default Rule Paradigm and the Limits of Contract Law*, 3 S. CAL. INTERDISC. L.J. 389, 416 (1993); Alan Schwartz, *Incomplete Contracts*, *in* 2 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW, *supra* note 5, at 277, 280 [hereinafter Schwartz, *Incomplete Contracts*]; *see also* Robert E. Scott, *The Case for Formalism in Relational Contract*, 94 NW. U. L. REV. 847, 875 (2000).

25. Schwartz, *Incomplete Contracts*, *supra* note 24, at 280.

26. *Id.* at 282.

27. *Cf.* Alan Schwartz, *Relational Contracts in the Courts: An Analysis of Incomplete Agreements and Judicial Strategies*, 21 J. LEGAL STUD. 271 (1992). In this earlier article, Schwartz argued that courts interpret contracts aggressively when bargaining defects exist and when the interpretation can be justified on the basis of verifiable information.

economics, but the appeal derives from the methodological decision to treat individuals as rational and courts as hampered by information asymmetries.[28]

D. *Unconscionability and Consumer Protection*

The premises of economics push in the direction of freedom of contract, and this current can be resisted only with difficulty. If parties are rational, they will enter contracts only when it is in their self-interest, and they will agree only to terms that make them better off. Courts that refused to enforce these terms would make it more difficult for future parties to use contracts to enhance their joint well-being. Therefore, courts should enforce the terms of the contract.

And yet courts do not always enforce the terms of contracts. They often refuse to enforce terms that seem harsh, oppressive, or improper: strict liquidated damages provisions, expansive security arrangements, alienation of the equity of redemption, restrictive arbitration provisions, broad covenants not to compete, wagers, choice-of-forum clauses and disclaimers of warranties in fine print or confusing language, and even price terms that seem too high or too low. Some of these practices derive from statutes (for example, usury laws); others arose in equity or the common law. The catchall term is unconscionability, but the relatively unusual application of this doctrine by courts only deflects attention from the widespread judicial scrutiny of transactions involving consumers, much of it in the form of interpretive presumptions that can interfere as much with freedom of contract as prohibitions do.

Economics has been better at deflating standard explanations for unconscionability and related doctrines than at explaining these doctrines. Let me say a few words about these standard explanations.

1. *Unequal Bargaining Power*

Courts sometimes say that a contract is unconscionable because of the unequal bargaining power of the seller and buyer. It is not always clear what courts mean when they use this term, but the closest economic concept is that of market or monopoly power. A seller has market power if it can increase the price of the good above its marginal cost by restricting supply. As is well known, such behavior is inefficient in the Kaldor-Hicks sense, and forcing the seller to sell at marginal cost would in theory eliminate a deadweight cost.

28. For a parallel argument, see *infra* Section II.B.

Nonetheless, economists typically argue that courts should not avoid contracts because of the unequal bargaining power of the parties. When contracts appear to have very high price terms, a court could determine only with great difficulty whether the high price is due to market power or fluctuations in the costs of inputs. A high interest rate, for example, could result from the creditor's judgment about the risk of default posed by a particular debtor, and generally courts should defer to such judgments. A determination that the creditor has market power requires an evaluation of the structure of the market, a notoriously difficult enterprise usually reserved for antitrust litigation. A seller or creditor with temporary market power as a result of a patent, or some innovation that other market participants have not had a chance to imitate, should (arguably) be permitted to reap above-market returns, for that is how innovation is encouraged in a market economy.

When contracts appear to have harsh nonprice terms, there is another reason for thinking that these terms are unobjectionable. Even if the seller or creditor has market power, it has the right incentive to supply the terms that parties desire. For example, a debtor might be willing to consent to a harsh remedial term in return for a low interest rate.[29] And a supplier might be willing to give the buyer the power to terminate the contract with little notice, if that is the only way to get the buyer's business. The party with market power will supply terms if the other parties want them and will charge them a fee, but will not force terms on parties that do not want them, for generally the most efficient way to exploit market power is through the price term.[30] Although there are models in which a combination of market power and asymmetric information can result in inefficient terms, they justify nonenforcement only under complex and hard-to-identify conditions.[31]

These theories do not describe what courts do. Courts permit the harshness of nonprice, and occasionally price, terms to influence them, and they seem to attach significance to unequal bargaining power. For this reason, most economic work is cast as a normative critique of the judicial practice.

29. Richard A. Epstein, *Unconscionability: A Critical Reappraisal*, 18 J.L. & ECON. 293, 305-06 (1975).

30. Alan Schwartz, *A Reexamination of Nonsubstantive Unconscionability*, 63 VA. L. REV. 1053, 1071-76 (1977); Alan Schwartz & Louis L. Wilde, *Imperfect Information in Markets for Contract Terms: The Examples of Warranties and Security Interests*, 69 VA. L. REV. 1387, 1458 (1983).

31. *See* Alan Schwartz & Louis L. Wilde, *Intervening in Markets on the Basis of Imperfect Information: A Legal and Economic Analysis*, 127 U. PA. L. REV. 630, 666-71 (1979). For a discussion of the literature in the context of consumer finance, see Richard Hynes & Eric A. Posner, *The Law and Economics of Consumer Finance*, 4 AM. L. & ECON. REV. 168 (2002).

2. *Lack of Information*

Courts sometimes say that a contract is unconscionable because one party, usually a consumer, lacks sophistication. Lack of sophistication is not the same thing as lack of information, but lack of information does seem to play a role in the cases. When terms are harsh and complex or hard to read, and consumers are unsophisticated, courts often express doubt that the consumers understood their obligations under the contract. This has led economists to investigate the role of information deficiencies in contract enforcement.

The topic is too complex to discuss here in any detail, but let me make a few observations. Consumers who lack information have incentives to acquire information. Some consumers will acquire information more easily than others; these are the people who read *Consumer Reports*. But the other consumers can free-ride on the efforts of the first group. If sellers cannot easily distinguish informed and uninformed consumers, they cannot exploit the latter by charging them a higher price. Thus, information deficiency alone does not justify judicial intervention.[32]

In addition, sellers have incentives to provide information to otherwise uninformed consumers. If seller X has lower costs than seller Y, and thus can charge lower prices and obtain a profit, X will invest in advertising in order to attract consumers from Y. There are limits, however, to the amount of information X will provide. If X's cars are cheaper than Y's cars, X has the right incentives; but if X knows that its cars in general are more dangerous than consumers believe, X has no incentive to provide that information.[33] Supplying such information is costly, both intrinsically and in the form of lost sales, and X does not internalize the benefits when he honestly warns of the dangers of automobile travel and consumers refrain from buying cars and avoid being injured.[34]

3. *Summary*

In sum, a simple model of the consumer-goods market implies that courts should not use the unconscionability doctrine to strike down contracts. More complex models suggest that courts should ignore bargaining power or should take it into account only under narrow conditions. Yet courts frequently criticize the inequality of bargaining

32. Schwartz & Wilde, *supra* note 30, at 1422-23.

33. *See* Howard Beales et al., *The Efficient Regulation of Consumer Information*, 24 J.L. & ECON. 491, 505-06 (1981).

34. X's incentives are suboptimal even if cars are safer, rather than more dangerous, than consumers think because X would not internalize gains to Y that would result if X revealed this information to consumers. Monopolists might gain more from information disclosure than competitors, but there are further complications. *See id.* at 507-08.

power between consumer and seller, imply that this fact may justify avoidance of the contract, and do not elaborate any further on the question of why bargaining power matters in some cases but not others. Other models suggest that courts might improve information asymmetries when consumers do not engage in enough comparison shopping or when competitive pressures do not force sellers to reveal information. Yet courts rarely pay attention to these factors when applying the unconscionability doctrine.

E. *Mistake*

In some circumstances, courts avoid contracts that are the result of mistakes. If the parties committed a mutual mistake as to a basic assumption of the contract, or if one party committed a mistake that the other party could have detected, the adversely affected party will sometimes have the right to avoid the contract.

In theory, parties could design contracts that released one or both parties who made a mistake. Consider a contract between Buyer and Seller for the sale of a cow.[35] Buyer and Seller might believe that the cow is barren when in fact she is fertile, in which case Seller will want to avoid the contract. Or Buyer and Seller might believe that the cow is healthy when in fact she is ill, in which case Buyer will want to avoid the contract. In either event, the parties can design the contract accordingly. The parties could enter a contract giving the Seller the right to withdraw from the contract if the cow proves to be fertile, or make performance contingent on subsequent confirmation that the cow is barren. And in the second case, Seller could give Buyer a warranty against illness or not, depending on how they want to allocate the risk. The general point is that if parties are rational, they know that they can make mistakes, and they will design the contract in a way that assigns this risk in the appropriate manner.

One might respond that because the parties are, by hypothesis, mistaken, it does not occur to them to build these contingencies into the contract. This is what courts mean when they say that the mistake was about a basic assumption of the contract. But rational parties always know that something could happen that makes performance more or less costly to Seller, and more or less valuable to Buyer. It could be that the cow has a hidden characteristic, good or bad; it could be that market conditions will change, so that a cow gains or loses value relative to other goods. From an economic perspective, there is nothing special about the cow being fertile or ill, nothing that distinguishes this contingency from a change in the price

35. Sherwood v. Walker, 33 N.W. 919 (Mich. 1887).

caused by a shift in market conditions. Parties can design contracts that take account of all these contingencies.

If this argument is correct, there is no reason for courts to release parties when one or both of them make mistakes. It would be like releasing an insurance company from a fire-insurance contract on the ground that the insurance company mistakenly believed that a fire would not occur. From an economic perspective, parties cannot make mistakes: They have probability distributions that reflect information they have about the world. They know that they do not possess the absolute truth and would not believe otherwise.[36]

In order to explain the mistake doctrines, then, we need to make additional assumptions. One possible assumption, which by now should be familiar, is that "transaction costs" prevent parties from designing optimal contracts. This is the implicit route taken by Rasmusen and Ayres in an article on the mistake doctrines.[37] Before we turn to their argument, we should observe that using this assumption makes the analysis of the mistake doctrine the same as the analysis of any problem of contractual interpretation, where, as law and economics assumes, transaction costs prevent parties from defining a crucial term, like "chicken" in the *Frigaliment* case. The parties do not make a "mistake" in the ordinary sense of the term; they rationally choose to leave a contract incomplete in light of the costs of completing it. If one thinks that courts should use majoritarian defaults to determine such terms, then one should think that majoritarian defaults should also determine the parties' obligations if the cow is fertile or ill. On this view, the mistake doctrine should also have the same remedial implications as contract-interpretation disputes—namely, enforcement of the judicial interpretation rather than rescission and restitution—but of course it does not.

To evaluate the mutual and unilateral mistake doctrines, Rasmusen and Ayres assume that parties can set a price but that they cannot make performance contingent on the occurrence of the desired states of the world (either directly or through the use of an optimal incomplete contract). It is in this sense that they operationalize the concept of mistake, taking their cue from the work on contract interpretation. Seller expects an average cost to perform, c, and Buyer expects an average valuation, v, such that $v > c$, on average, but in some states of the world $v < c$, and trade should not occur. Ayres and Rasmusen investigate the question whether the mistake doctrine should release the parties from the contract (presumably, at the request of

36. Eric Rasmusen & Ian Ayres, *Mutual and Unilateral Mistake in Contract Law*, 22 J. LEGAL STUD. 309, 315 & n.13 (1993).

37. *Id.* at 315.

Buyer if the price is higher than his realized valuation, or of Seller if the price is lower than her realized cost).

Their analysis is complex, so let me focus on their conclusions for mutual mistake. The doctrine, by excusing performance when both parties are uninformed or mistaken, would seem to create incentives for parties to avoid performance and to gather information. The question is whether these incentives are efficient.

One possible advantage of the mutual mistake doctrine is that it enables the parties to avoid a second transaction in order to reverse the initial contract when it turns out that $v < c$. But the reversal will occur only when both parties are also uninformed; if one party is informed, then the other party cannot avoid an inefficient contract by claiming mutual mistake. Further, the doctrine enables either party to avoid the contract when (despite the mistake) $v > c$: Buyer will avoid the contract when the price $p > v$, and Seller will avoid the contract when $p < c$.[38] If you permit parties to breach without paying damages, then inefficient breach will result.

The mutual mistake doctrine also affects the parties' incentives to gather information prior to entering contracts. But it does not affect the incentives in a desirable way. Suppose that parties can acquire information about the value of the good at some cost. If the acquisition of information does not increase the value of the good, and the cost of acquiring information is high enough, the mutual mistake doctrine properly encourages the parties to remain ignorant but also results in too many rescissions. When the cost of information is lower, the mutual mistake doctrine encourages each party to acquire information in order to prevent the other party from invoking mutual mistake, even though the additional information does not increase the value of the good. On the other hand, if the acquisition of information *does* increase the value of the good, then the mutual mistake doctrine gives too little incentive to acquire information, compared to a rule of enforcement.[39]

The mutual mistake doctrine is hard to reconcile with economic premises, both because the doctrine by its terms appeals to cognitive errors excluded from economic analysis, and because the doctrine encourages behavior that would be suboptimal if people did not make those errors, as economics assumes.[40] But even putting aside cognitive errors, Ayres and

38. *Id.* at 320.

39. *Id.* at 331-32.

40. Indeed, most contracts scholarship on information asymmetries assumes that the parties know about their information advantage or disadvantage and make a strategic decision to reveal information, demand a price adjustment, refuse to enter a contract, and so forth. The articles usually focus on the proper damages remedy given these assumptions. In this context, Rasmusen and Ayres's argument can be understood as an investigation of the conditions under which a "zero damages" rule would be superior to expectation damages and alternative rules. For examples of this literature, see Richard Craswell, *Performance, Reliance, and One-Sided Information*, 18 J. LEGAL STUD. 365 (1989) (analyzing the effect of damages rules on the promisor's incentive to disclose private information about the probability of performance); Richard Craswell, *Precontractual Investigation as an Optimal Precaution Problem*, 17 J. LEGAL STUD. 401 (1988) (analyzing the effect of damages rules on the parties' incentives to acquire information, prior to contracting, about their ability to perform); Peter A. Diamond & Eric Maskin, *An Equilibrium Analysis of Search and Breach of Contract, I: Steady States*, 10 BELL J. ECON. 282 (1979) (analyzing the effect of damages rules on parties' incentives to search for contract partners); and Steven Shavell, *Acquisition and Disclosure of Information Prior to Sale*, 25 RAND J. ECON. 20 (1994) (analyzing the effect of disclosure rules on the incentive to acquire and disclose information).

Rasmusen do not offer a clear replacement for the mutual mistake doctrine. They show that other doctrines, such as no excuse and unilateral mistake, dominate mutual mistake in different contexts, but not that one of these doctrines or some alternative would be optimal in general, nor that a court could distinguish the conditions—whether, for example, acquisition of information would increase the value of a good or not—under which the different rules have advantages.

F. *Impossibility*

Courts sometimes release promisors from performance when performance is "impossible" or "impracticable." Posner and Rosenfield argue that these doctrines efficiently shift risk from the promisor when the promisor is more risk-averse than the promisee.[41] Suppose a seller cannot insure itself against a strike by its workforce, but that the buyer can easily arrange for deliveries from alternative sellers if supply from the first is cut off. If the seller subsequently cannot make deliveries because of a strike, a court might excuse the seller from its contractual obligations on the grounds of impossibility, with the real reason being that buyer could have insured against this contingency more easily than the seller could have.

Subsequent work casts doubt on this argument. First, Posner and Rosenfield's argument neglects the other incentives of the parties. If the seller pays no damages or a limited amount like restitution, it has no incentive to perform when it is efficient to do so. The argument assumes that the court can determine whether the cost of performance exceeds its value to the buyer. But in other contexts, the justification of expectation damages, for example, it is assumed that the court cannot make this determination.[42]

Second, and more important, the impossibility and impracticability doctrines do not spread risk in the efficient way. To see why, imagine a risk-averse seller and a risk-neutral buyer. The optimal resolution of a

41. Richard A. Posner & Andrew M. Rosenfield, *Impossibility and Related Doctrines in Contract Law: An Economic Analysis*, 6 J. LEGAL STUD. 83 (1977).

42. *See* Michelle J. White, *Contract Breach and Contract Discharge Due to Impossibility: A Unified Theory*, 17 J. LEGAL STUD. 353, 360 (1988).

dispute will place all the risk on Buyer; in effect, the court writes an insurance policy of which Seller is the beneficiary. Such an insurance policy would give Seller the same payoff in both states of the world—the breach state and the performance state. The impossibility and impracticability doctrines, however, do no such thing. They, at best, give Seller a zero payoff in the breach state (and less if Seller has incurred some costs or must make restitution), instead of giving Seller some amount between zero and its profits. Indeed, the risk-sharing argument implies that in some cases Seller should pay negative damages, something which is, of course, never observed.[43]

The excuse doctrines are hard to understand from the economic perspective. Sophisticated parties know that contingencies might occur that will make performance impossible or extremely costly. If they want to share the risk of these contingencies, they can write excuses into the contract. We observe this behavior not just in the use of force majeure clauses; excuses are frequently built into the central terms of the contract. Insurance contracts contain exclusions; ordinary sales contracts shift risk by tying the price to market indices. Firms can buy general insurance policies, or self-insure, and are usually risk-neutral with respect to run-of-the-mill contracts. It might be true that the cost of describing the parties' obligations prevents parties from assigning all the risks, but it remains doubtful that courts have the information necessary to repair the insurance market.

G. *Consideration and Promissory Estoppel*

Economics assumes that people exchange promises when both benefit from the exchange, but it does not follow that the law should enforce all promises. Courts make errors, and legal sanctions are sometimes clumsier than nonlegal sanctions. As a result, people who make and receive promises often do not expect, and would not want, courts to provide legal remedies if the promisor breaks the promise. But when the promisor wants the promise to be legally enforceable, and the promisee expects the promise to be legally enforceable, courts should enforce promises.[44]

43. *See* Alan O. Sykes, *The Doctrine of Commercial Impracticability in a Second-Best World*, 19 J. LEGAL STUD. 43, 48 (1990); White, *supra* note 42, at 375; *see also* Victor P. Goldberg, *Impossibility and Related Excuses*, 144 J. INSTITUTIONAL & THEORETICAL ECON. 100 (1988) (expressing skepticism about the risk-sharing aspect); Polinsky, *supra* note 7 (arguing that endorsement of liquidated damages is the optimal remedy for risk sharing); George G. Triantis, *Contractual Allocations of Unknown Risks: A Critique of the Doctrine of Commercial Impracticability*, 42 U. TORONTO L.J. 450 (1992) (criticizing the excuse doctrines).

44. This conclusion excludes the possibility of a market failure. One might also stipulate a de minimis requirement: Courts should not enforce promises when the cost to the legal system exceeds the gains to the parties. Charging the litigants a fee would be a better response.

Economics, then, implies that courts should enforce promises when parties want their promises to be enforceable, and not otherwise. Consistent with this view, courts both routinely enforce promises and respect terms of agreements that disclaim legal enforceability.[45]

But these simple ideas do not explain the main doctrines that draw a line between the legally enforceable promise and the unenforceable promise, namely, the consideration and promissory estoppel doctrines.

The consideration doctrine holds that a court cannot enforce a promise if it was not exchanged for "consideration," a legal benefit to the promisor or detriment to the promisee (or, in the modern formulation, a promise or performance that was bargained for). In essence, the doctrine knocks out of court promises that are not part of a quid pro quo. Such promises include option contracts, promises to give a gift, and open-ended agreements that bind one party but not the other.

Yet these promises are unobjectionable from an economic perspective. An option contract—for example, a promise to keep open an offer to sell something while the offeree investigates its value—might be the only way to attract the interest of a prospective purchaser. A promise to give a gift enables the promisee to rely in anticipation of receiving the benefit and enables the promisor to defer performance until the funds or goods are acquired. Open-ended contracts—where, for example, one side commits itself to purchase goods produced by the other side—are often efficient methods for shifting risk, with the legally unconstrained party bound by reputational concerns and nonlegal sanctions.[46]

The courts, possibly because they recognize the force of these arguments, have whittled down the consideration doctrine. Its main function is now to deny enforcement of promises to give gifts.[47] The consideration doctrine also serves, under the *Restatement*, as a formality: Options are unenforceable unless the parties "recite" consideration.[48] But there is no reason to require parties to recite a consideration as opposed to reciting that they want their option to be enforceable. The same can be said for Holmes's argument that the consideration doctrine was always just a formality, so gift promises would be enforced if the promisee gave nominal consideration to the promisor.[49] Efforts to rationalize this practice as a way

45. *See, e.g.*, Empro Mfg. Co. v. Ball-Co Mfg., 870 F.2d 423 (7th Cir. 1989).

46. *See* Charles Goetz & Robert E. Scott, *Principles of Relational Contracts*, 67 VA. L. REV. 1089, 1149-50 (1981).

47. The exception for charitable gifts only complicates the puzzle. If promises to give gifts are socially desirable, then the exception makes sense, but the general unenforceability of gift promises does not; if promises to give gifts are not socially desirable, then the exception does not make sense.

48. RESTATEMENT (SECOND) OF CONTRACTS § 87 (1981) [hereinafter RESTATEMENT].

49. OLIVER WENDELL HOLMES, THE COMMON LAW 294-95 (Little, Brown & Co. 1946) (1881); *see also* Lon L. Fuller, *Consideration and Form*, 41 COLUM. L. REV. 799 (1941) (criticizing Holmes's argument).

of ensuring that courts can distinguish enforceable and unenforceable promises fail because they do not explain the "form" of the formality.

The doctrine of promissory estoppel might seem consistent with economics because it does not forbid courts to enforce value-enhancing gratuitous promises. The doctrine does place a limit on the enforcement of promises, however, and that limit is the requirement of promisee reliance. This limit is not consistent with economics. If a person wants to make a gratuitous promise, it must be because he wants to make the promisee better off. The promisor can make the promisee better off regardless of whether the promisee relies on the promise or can be proved to have done so. Economic analysis therefore suggests that enforceability of a promise should not depend on whether the promisee relied, or relied reasonably.[50]

Promissory estoppel can also be understood as a device for relaxing the consideration doctrine's prohibition on liability for precontractual reliance. Parties often spend some time negotiating the terms of a contract before entering it; frequently, one or both parties will make investments during this period in anticipation of the eventual success of the negotiations. An example is the relationship between a franchisor and a franchise applicant, which can extend for months or years before the granting of the franchise.[51] During this time, the franchisor might require the applicant to acquire experience as an employee in another franchise business, or undergo training. To induce the applicant to make these investments of time and effort, the franchisor might make vague or contingent promises that the franchise will be awarded. Even when these promises are not definite enough to form contractual commitments, applicants who are not awarded the contract can sometimes obtain damages for their reliance costs, on the basis of promissory estoppel.

Several scholars have considered the possibility that promissory estoppel is efficient because it protects the promisee's investment.[52] This view is at first sight attractive because the promisor's behavior, when not justified by the discovery that the promisee is unfit, seems opportunistic. In theory, the promisor can hold up the promisee after the promisee has invested and demand from the promisee additional fees or obligations that extract all the surplus generated from the promisee's investment.

But if courts could reliably verify the promisee's behavior, and thus distinguish the promisee who proves merely to be unfit and the promisee who is the victim of holdup, then the parties could enter a contract at the

50. *See* Steven Shavell, *An Economic Analysis of Altruism and Deferred Gifts*, 20 J. LEGAL STUD. 401, 420 & n.24 (1991).

51. *E.g.*, Hoffman v. Red Owl Stores, 133 N.W.2d 267 (Wis. 1965).

52. Richard Craswell, *Offer, Acceptance, and Efficient Reliance*, 48 STAN. L. REV. 481, 489-95 (1996); Avery Katz, *When Should an Offer Stick?: The Economics of Promissory Estoppel in Preliminary Negotiations*, 105 YALE L.J. 1249, 1270-77 (1996).

beginning of their relationship, one that specified what the promisee must invest and how the promisee will be evaluated. As far as I know, parties do not enter such contracts, presumably because they do not believe that courts can make these distinctions.[53] But if courts cannot, then the use of promissory estoppel to protect reliance is not justified.

In addition, an efficient promissory estoppel doctrine would not require courts to compensate all of the promisee's reliance. If it did so, promisees would overinvest in reliance. Courts would need to determine how much reliance is efficient in each case, and then award damages only equal to efficient reliance, undercompensating parties that rely too much, or not compensating them at all. The proper award would depend on such factors as the cost of, and return on, investment; the probability that the preliminary relationship would yield a franchise;[54] and the parties' incentives to reveal information to each other.[55]

Craswell studied a group of cases in which equitable estoppel or promissory estoppel arguments were advanced by offerees in order to prevent an offeror from withdrawing an offer, and found that courts were more likely to rule in favor of the offeree when reliance on the offer is efficient.[56] But he disclaims any intention to show that the outcomes of the cases are themselves efficient, for just the reasons given above: "[T]here are several factors other than the efficiency of [the offeree's] reliance that can affect the desirability of a commitment."[57] The methodological difficulty of showing that contract doctrine is efficient dissuades Craswell from making the attempt.

H. *Summary: Descriptive Versus Normative Failure*

The charge of descriptive failure will not surprise scholars familiar with the literature on economic analysis of contract law. The inefficiency of contract law is a theme of Shavell, Goetz and Scott, and Schwartz on expectation damages; Epstein and Schwartz on the unconscionability doctrine; Ayres and Rasmusen on the mistake doctrine; Sykes and White on

53. In response to the growth of precontractual liability, franchisors now require franchise applicants to sign waivers.

54. *See* Craswell, *supra* note 52, at 499-501; *see also* Lucian Arye Bebchuk & Omri Ben-Shahar, *Precontractual Liability*, 30 J. LEGAL STUD. 423 (2001); Jason Scott Johnston, *Communication and Courtship: Cheap Talk Economics and the Law of Contract Formation*, 85 VA. L. REV. 385 (1999); Katz, *supra* note 52.

55. Avery Katz, *Contract Formation and Interpretation*, *in* 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW, *supra* note 5, at 425, 429-30.

56. Craswell, *supra* note 52, at 531-36.

57. *Id.* at 507.

the impossibility doctrine; Craswell on promissory estoppel; and others.[58] None of the authors goes so far as to deny that contract law is efficient. Each examines only a small slice of contract law, and normal methodological practice cautions against making exaggerated claims. But for the observer who looks at the steady accumulation of failures over thirty years, the conclusion is inescapable.

Some readers might agree with Ayres and Craswell that since most law-and-economics authors do not claim that they explain doctrine, and instead make normative recommendations, it is not worth making too much of this failure. Perhaps not, but I do not think that the distinction between descriptive and normative scholarship is so clear. The doctrinal structure of contract law exerts force on the scholarly analysis. That is why so many authors try to rationalize the doctrine, or propose incremental changes, rather than coming to the austere conclusions of Schwartz and others influenced by the incomplete contracts perspective. Courts and legislatures are more likely to pay attention to scholarly recommendations that follow naturally from the logic of contract law, than those that float down from the ether, for courts and legislatures have no good reason—no economic crisis, no foreign contract-law system that is clearly superior, no chorus of complaints like those heard about the tort system—to think that there is anything wrong with the system of contract law that we have. When economics was able to keep the descriptive and normative together, when it was able to say that contract law was essentially efficient but for some tweaking here and there, it had the potential to influence decisionmakers, for it worked with the past, not against it, and did not force decisionmakers to reject the past on the basis of conjectures founded on empirical postulates that could not be verified.

The descriptive failure of the models takes two forms. Simple models, which examine only one or two margins of contractual behavior, fail to predict contract law as it exists. The other models are more complex because they examine a greater variety of behavior, or because they rely on more complicated ideas, such as information asymmetry. These models sometimes fail because they make predictions that are inconsistent with contract law. But more often they fail because they are indeterminate. The models incorporate variables that cannot be measured, and to which one cannot with any confidence attach general ranges or distributions.

To repeat one example, recall that the choice between the *Hadley* rule and the rule of expansive liability depends, among many other things, on the shape of the distribution of buyer valuations. If, in terms of numbers of

58. *E.g.*, Avery Katz, *Transaction Costs and the Legal Mechanics of Exchange: When Should Silence in the Face of an Offer Be Construed as Acceptance?*, 9 J.L. ECON. & ORG. 77, 90 (1993) (finding acceptance-by-silence doctrines' efficiency to be "imperfect at best").

buyers or the magnitudes of their valuations, the distribution is lopsided toward low valuations, then the *Hadley* rule is more likely to be superior. No one has tried to determine the shape of this distribution through empirical research, and indeed it is hard to imagine how this could be done. It is also foreign to the fact-finding activities of courts and legislatures. (One must also fix the relevant population, take a stab at guessing the cost of communicating, and so forth.) Accordingly, it might be best to assume that the distribution is uniform or normal, in which case neither rule is superior.

A further point is that the descriptive approach has not been fruitful in the way that it is in other areas of economics. In these other areas, the thing to be explained is always partly hidden. It makes sense to develop a hypothesis because in the process of testing it one learns new things about the world, resulting in a productive dialectic between theory and data. By contrast, the thing to be explained by the economic analysis of contract law—contract doctrine—is known, or thought to be known. And although at one time some scholars thought that outcomes of cases might diverge from contract doctrine, with the outcomes reflecting efficiency—in which case, generating and testing hypotheses would make sense—today this view has few adherents.[59] Judges have no reason to describe doctrine in a way that misrepresents the outcomes of cases.

Rather than arguing that their models explain contract doctrine, most authors argue that their models can be used to criticize or defend contract doctrine. But the normative weaknesses of their models follow as a matter of course. Simple models do not justify legal reform because these models exclude relevant variables. Complex models do not justify legal reform because the optimal rule depends on empirical conditions that cannot be observed.[60]

One might respond that even if economic models cannot generate a determinate optimal contract law, they helpfully identify the costs and benefits of different legal rules. Before the economic analysis of expectation damages and specific performance, a court trying to decide whether to push the doctrine in one direction or the other had little to go on. Economic analysis identified factors of which judges should take account, factors that include the cost of renegotiation and the advantages of permitting breach. Even if economic analysis cannot determine the magnitude of these costs and benefits, and the extent to which they offset or

59. There has, however, been a recent effort to defend this proposition. *See* Fred S. McChesney, *Tortious Interference with Contract Versus "Efficient" Breach: Theory and Empirical Evidence*, 28 J. LEGAL STUD. 131, 176-84 (1999).

60. Others have made similar observations about normative law and economics. *See, e.g.*, Lewis A. Kornhauser, *Constrained Optimization: Corporate Law and the Maximization of Social Welfare*, *in* THE JURISPRUDENTIAL FOUNDATIONS OF CORPORATE AND COMMERCIAL LAW, *supra* note 3, at 87, 112-13.

interact with each other, the judge who knows about them is more likely to make a wise decision than a judge who does not.

This defense has an air of plausibility but also distressingly open-ended and unambitious implications. The last decade has witnessed a piling on of relevant factors, but no increasing clarity about the function of contract law, and a wise judge might, in order to avoid paralysis, simply ignore them. But the scholarship itself is mute about its own weaknesses. Part III will provide some methodological reasons for skepticism. Before we get there, however, we can gain additional insight by examining the literature on incomplete contracts.

II. THE THEORY OF INCOMPLETE CONTRACTS

The literature on incomplete contracts diverges from the law-and-economics literature, though they overlap in many ways. The theory of incomplete contracts was motivated primarily by descriptive curiosity about the nature of private contracting, not about contract law. As a result, contract law is usually treated in an exceptionally simple manner, as a system that specifically enforces contractual terms when the underlying behavior can be verified by courts.[61] This assumption enables scholars to focus on the parties' choice of contractual form. By contrast, law and economics generally assumes that parties choose simple contracts—contracts with a fixed price and quantity and sometimes a liquidated damages clause—and focuses on the effect of different legal rules on contractual behavior.

The incomplete contracts literature poses the following question to law and economics: Why would rational parties choose noncontingent contracts when more sophisticated contracts would enable parties to achieve better results? And if parties did choose more sophisticated contracts, why would courts need to do anything other than enforce the terms of these contracts? If courts only enforced the terms of contracts, much of contract doctrine, and much of the law-and-economics literature, would be irrelevant.

The following discussion of the theory of incomplete contracts serves two purposes. First, it allows us to examine whether the descriptive failure of law and economics is the result of economic methodology in general, or

61. *See, e.g.*, OLIVER E. WILLIAMSON, MARKETS AND HIERARCHIES: ANALYSIS AND ANTITRUST IMPLICATIONS (1975); Oliver Hart & John Moore, *Incomplete Contracts and Renegotiation*, 56 ECONOMETRICA 755 (1988). *Incomplete Contracts and Renegotiation* is generally considered the seminal article, though the literature has roots in Williamson's work. My reference to the literature on "incomplete contracts" is intended to encompass articles that are formally based on complete contract models but are concerned with the question of why contracts are incomplete, and not just articles that include transaction costs as an element in the model.

of the law-and-economics approach in particular. Second, it sheds light on the methodological difficulties hidden in the concept of transaction costs.

A. *Premises and Basic Results*

The incomplete contracts literature focuses on two of the kinds of incentives we have been discussing: the incentive to invest (or "rely") and the incentive to perform or breach. An efficient or "first best" contract does two things: It ensures that (say) Seller performs when her cost is less than Buyer's valuation, and not otherwise ("efficient trade"), and it ensures that Buyer (and/or Seller) invests the right amount ("efficient investment").

As we have seen, there is a tension between efficient trade and efficient investment. A simple way of ensuring efficient trade requires Seller to pay Buyer's valuation if she does not perform. For example, under the rule of expectation damages, if Seller fails to perform, she must pay Buyer's valuation. Thus, she performs if and only if her cost is less than Buyer's valuation, the condition for efficient trade. However, Buyer will expect to receive his valuation whether performance occurs (in which case he gets the good) or not (in which case he gets damages equal to his valuation). Expecting to receive his valuation in both states of the world, Buyer will overinvest.[62]

All of this should be familiar from our discussion of the law and economics of remedies. The difference between the two literatures is in the next step. Where law and economics evaluates alternative legal rules according to their impact on the efficiency of contractual behavior, the incomplete contracts literature analyzes how the parties might design their contract in order to achieve efficiency. If Seller and Buyer are rational, they will want to prevent Buyer's overinvestment, with the parties sharing the surplus that comes from eliminating this inefficiency. They can do so through correct contractual design, assuming courts will specifically enforce it.[63]

The simplest solution would be to write a contract that says that the parties must trade if $v > c$, and that Buyer must make optimal investment r. If the court could observe the valuations and the investment, then it could use specific performance or a penalty in order to force the parties to engage in efficient behavior. But then we would have a complete contract, and such contracts, the argument goes, are never used. In fact, contracts are

62. *See supra* text accompanying notes 11-13.

63. A few authors have also examined rules such as expectation damages, but they produce optimal incentives only under narrow conditions. *See, e.g.*, Aaron S. Edlin & Stefan Reichelstein, *Holdups, Standard Breach Remedies, and Optimal Investment*, 86 AM. ECON. REV. 478, 495 (1996).

incomplete, because transaction costs prevent the parties from putting all relevant, that is, value-maximizing, obligations in the contract.

The literature stipulates that transaction costs mean that the investment is not verifiable by a court, so the parties gain nothing by putting the optimal investment in the contract. For various reasons,[64] this assumption is thought to be a more satisfactory way of capturing the concept of transaction costs than, say, stipulating that there is a cost to writing an obligation down or that the parties must write fixed price or noncontingent contracts—the two preferred strategies in the law-and-economics literature. In any event, arguments about damages rules in the law-and-economics literature probably do not turn much on exactly how transaction costs are characterized.[65]

Even if the investment is nonverifiable, the parties could design a contract that provides efficient incentives. The contract would give Buyer the right to set the price at the time of performance and make an offer to Seller.[66] Seller would have the right to reject the transaction (and receive liquidated damages, assumed for the sake of the example to be zero) or to accept the transaction and accept the price announced by Buyer, in which case Buyer must accept delivery at that price.

This contract would achieve first-best efficiency. To see why, suppose first that $v > c$. Buyer will set the price equal to (or slightly higher than) Seller's cost, which Buyer observes. If Buyer set the price lower than Seller's cost, Seller would refuse to trade, and Buyer would gain nothing. But Buyer has no reason to set the price higher than Seller's cost, which would only reduce Buyer's own return. Thus, Buyer sets the price equal to c. But if $v < c$, Buyer will set the price at some low level in order to prevent Seller from demanding trade, for Buyer does not want to receive less than he pays. Seller will thus trade if and only if $v > c$, so the conditions for efficient trade are met.

When Buyer makes his investment, he knows that he will receive the goods only if $v > c$, and not otherwise. Thus, Buyer makes his investment with an eye to obtaining a return only in the performance state of the world. Indeed, Buyer will obtain the full residual of the investment because he will

64. *See, e.g.*, Ilya Segal, *Complexity and Renegotiation: A Foundation for Incomplete Contracts*, 66 REV. ECON. STUD. 57, 72-73 (1999).

65. The assumption that investments are not verifiable would not, for example, undermine Shavell's arguments about the relative efficiency properties of expectation and reliance damages, *see* Shavell, *supra* note 11, at 147, but would undermine Cooter's argument that the optimal damage measure is the amount that would compensate the victim if efficient investment had occurred, *see* Cooter, *supra* note 12, at 14.

66. I follow the discussion in Benjamin E. Hermalin & Michael L. Katz, *Judicial Modification of Contracts Between Sophisticated Parties: A More Complete View of Incomplete Contracts and Their Breach*, 9 J.L. ECON. & ORG. 230 (1993).

set price equal to c. Thus, Buyer has the incentive to engage in efficient investment.

The contract's trick, in this case, is to give the party with the investment decision the residual from trade. As a result, the party has the correct incentives both to trade and to invest. The other party, Seller, must of course be compensated for her expected costs, and an ex ante transfer from Buyer to Seller accomplishes this task.

This is only the first step in a literature that has become very lengthy and complicated. Authors have discussed such problems as two-sided investment, where Buyer can increase his valuation and Seller can reduce her cost;[67] cooperative investment, where Buyer can reduce Seller's cost and Seller can increase Buyer's valuation;[68] third party effects;[69] and so forth.

The most interesting thing about these models is that they predict that contracts will contain descriptions not of "physical" contingencies but of the bargaining procedures that parties must follow at the time of performance. Lawyers think of contracts as either providing absolute obligations (Seller must deliver widgets by December 1st) or conditional obligations, with the conditions referring to events that occur in the world such as a strike or price change (Seller must deliver widgets unless Seller experiences labor difficulties, etc.). In models of incomplete contracts, the bargaining procedures specified in the predicted contracts are designed to force parties to divulge, and act efficiently on the basis of, their realized valuations, Seller's cost, and Buyer's value, and so references to events in the world are unnecessary. If Seller suffers a strike, for example, and her costs rise above Buyer's valuation, Seller will exercise an option to pay money rather than produce and deliver the goods. The contract does not need to refer explicitly to Seller's obligations in case of a strike. Because the parties can foresee that their valuations might change, and can design bargaining procedures that elicit efficient behavior (or behavior no more inefficient than that which would occur under a simpler fixed price contract), they do not have to write down countless contingencies in their contract. For this reason, the guiding premise of law and economics, that

67. Philippe Aghion et al., *Renegotiation Design with Unverifiable Information*, 62 ECONOMETRICA 257 (1994); Tai-Yeong Chung, *Incomplete Contracts, Specific Investments, and Risk Sharing*, 58 REV. ECON. STUD. 1031 (1991). For a survey, see Klaus M. Schmidt, *Contract Renegotiation and Option Contracts*, *in* 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW, *supra* note 5, at 432.

68. Yeon-Koo Che & Donald B. Hausch, *Cooperative Investments and the Value of Contracting*, 89 AM. ECON. REV. 125 (1999).

69. *See infra* Section II.C.

transaction costs render contracts incomplete and justify court provision of default rules, seems too strong.[70]

And yet the incomplete contracts literature does not provide a promising alternative. The contracts that the models predict do not exist in the world. Instead, we see simple fixed price contracts or contracts that are conditional on a relatively small number of real world contingencies. Intuitively, the problem with the predicted contracts is that they are too complex for parties to design.[71] To write such contracts, parties would need to imagine their bargaining position if a breach should occur, and then work their way via backward induction to the optimal terms of the contract. People are not very good at backward induction. Yet the rationality assumptions of economics hold that they can do it perfectly.[72] This problem has led to some discussion among economists about whether a theory of contracts can avoid relying on a model of bounded rationality,[73] an issue to which I will return in Part IV.

B. *Freedom of Contract and Asymmetric Information: The Penalty Doctrine*

The incomplete contracts literature was motivated by the desire to explain contracting, not contract law; it is a branch of industrial organization, not of law and economics. But authors writing in this tradition have tried to explain some contract doctrines, and their efforts are worth examining because they shed light on the law-and-economics literature.

Hermalin and Katz show that as long as parties are symmetrically informed, courts cannot increase welfare by modifying, or refusing to enforce, contractual terms.[74] The logic should be familiar by now, and is indeed identical to longstanding defenses of freedom of contract. Parties have more information than courts about their preferences, and even if courts can obtain superior information ex post, at the time of performance or dispute, the parties will anticipate this behavior and design their contracts

70. For similar skepticism, on methodological and empirical grounds, see Schwartz, *Incomplete Contracts*, *supra* note 24. For example, if transaction costs were the cost of describing obligations for future states of the world, we would probably observe more complete contracts than we do. *Id.* at 277.

71. For a discussion, see Karen Eggleston, Eric A. Posner & Richard Zeckhauser, *The Design and Interpretation of Contracts: Why Complexity Matters*, 95 NW. U. L. REV. 91, 122-25 (2000).

72. *See* George J. Mailath, *Do People Play Nash Equilibrium? Lessons from Evolutionary Game Theory*, 36 J. ECON. LITERATURE 1347 (1998).

73. Oliver Hart, *Is "Bounded Rationality" an Important Element of a Theory of Institutions?*, 146 J. INSTITUTIONAL & THEORETICAL ECON. 696, 700-01 (1990); Oliver Hart & John Moore, *Foundations of Incomplete Contracts*, 66 REV. ECON. STUD. 115, 134-35 (1999); Eric Maskin & Jean Tirole, *Unforeseen Contingencies and Incomplete Contracts*, 66 REV. ECON. STUD. 83, 107-08 (1999); Segal, *supra* note 64, at 74.

74. Hermalin & Katz, *supra* note 66, at 245.

accordingly. Parties might, for example, leave the price term blank, to be filled in by the court ex post. But parties would not want a court to change the price term, or any other contractual term, based on its own judgment about what is ex post efficient.

If Hermalin and Katz are correct, then the instances where contract law does authorize courts to interfere with contract terms become a puzzle. The penalty doctrine is just one example, as we saw above, and indeed Hermalin and Katz criticize the penalty doctrine for the usual reasons.

Hermalin and Katz's argument is based on the assumption that the parties are symmetrically informed when they enter the contract. This assumption is not always true, and they acknowledge that if the parties are asymmetrically informed, judicial restrictions on contracts could increase welfare. Indeed, Hermalin made just such an argument with another coauthor three years earlier.[75] In the later article, however, Hermalin backs away from the asymmetry information argument.

In the earlier article, Aghion and Hermalin show that when the parties are asymmetrically informed, "legal restrictions on private contracts can enhance efficiency," as their title puts it.[76] The argument is best made by example. Imagine that a contractor has private information about the likelihood that it will perform a project on time. There are two types of contractors: The "good" type is more likely to perform on time than the "bad" type is. Buyers prefer good types to bad types, and so good types want to distinguish themselves from bad types. They do so by offering to pay an extremely high late fee or penalty if the performance is late. The bad type might mimic this signal, or not, but in either event the equilibrium, in which one or both types agree to the penalty, can be inferior to an equilibrium in which courts refuse to enforce penalties so that neither party can credibly agree to them.[77]

If we take this argument seriously, we should apply it not only to remedial terms. The same logic applies to the price term or, indeed, any other term of a contract. The contracts that emerge as a result of asymmetric information are simply inefficient contracts—it's not just that one term is inefficient, and the rest of the contract is efficient once that inefficient term is severed—and courts should refuse to enforce them even when a penalty clause is not activated by a breach.

To see why, imagine an employer who prefers educated applicants not because she cares about their education but because she believes that people

75. Philippe Aghion & Benjamin Hermalin, *Legal Restrictions on Private Contracts Can Enhance Efficiency*, 6 J.L. ECON. & ORG. 381 (1990).

76. *Id.* at 381.

77. The conclusion depends on the parameters of the model. Under some parameters, the separation of the types is superior to pooling; under other parameters, the opposite is the case. Aghion and Hermalin's point is that an inefficient equilibrium without judicial interference is possible, not certain. *See id.* at 399.

who graduate from college work harder than those who do not. The employer offers two employment packages, a low salary for those without diplomas, and a high salary for those with diplomas. Because a potential job applicant's decision to obtain an education has an external effect—it increases or reduces the employer's information about the quality of other potential applicants—the resulting equilibrium could involve inefficient signaling. For that reason, courts or legislatures might want to prohibit the employer's discriminatory behavior.[78] The Aghion and Hermalin argument implies that courts should scrutinize all contracts for inefficiency, and not just liquidated damages terms.[79]

Aghion and Hermalin, then, cannot distinguish the common law's treatment of remedial terms and non-remedial terms such as price terms. As a descriptive theory, it is a failure. As a normative theory, it is also not successful, as it assumes that courts have sufficient information to distinguish signaling equilibria where judicial intervention will increase welfare, and other equilibria where it will not. It is for this reason that Hermalin and Katz back away from the conclusions of Aghion and Hermalin. The former article, as I noted, expresses skepticism about the ability of courts to improve on parties' contracts even when asymmetric information is present. This is another descriptive failure because Hermalin and Katz cannot explain judicial restrictions on remedial terms.

C. *Freedom of Contract and Externalities: The Penalty Doctrine Again*

When two parties design a contract, they will choose terms that are optimal for themselves; they will not take account of the interests of third parties who might be affected by the contract. But there are such third parties. Consider a contract in which Seller must pay Buyer liquidated damages if Seller breaches. If liquidated damages are set very high, they might interfere with the effort of a third party (TP) to purchase the good from Seller, even though TP might value the good more than Buyer does.

To understand why, imagine that different potential TPs have different valuations. Among those who value the good more than Buyer does, some value it slightly more and some value it considerably more. When Seller and Buyer agree to relatively high liquidated damages, the latter clause prevents the low-value TPs from buying the goods (Seller won't sell to

78. The diploma example is taken from A. MICHAEL SPENCE, MARKET SIGNALING 14-15 (1974).

79. I am puzzled by Ayres's response to this argument. Ian Ayres, *Valuing Modern Contract Scholarship*, 112 YALE L.J. 881, 889-90 (2003). Information asymmetry is the standard justification for regulation of the insurance market, and regulation always involves the imposition of mandatory terms and other restrictions. Aghion and Hermalin themselves use the example of mandatory employment benefits to illustrate their argument. Aghion & Hermalin, *supra* note 76, at 401-03. I use the diploma example only because it is famous.

them because she has to pay high liquidated damages if she does), but it also enables Seller to extract a very high price from the high-value TPs, who must pay an amount at least as high as the already high liquidated damages. Under plausible conditions, Buyer and Seller jointly gain more by extracting the surplus from the high-value TPs than they lose by failing to sell to the low-value TPs. But this is inefficient, and the law should deter such behavior by refusing to enforce high liquidated damages provisions.[80]

This inefficiency disappears if renegotiation is possible: Ex post, the three parties will renegotiate so that TP will end up with the good, and efficient trade is achieved. But then the inefficiency shifts to the parties' investment incentives. The parties will choose inefficiently high liquidated damages to improve Seller's bargaining position vis-à-vis TP. Seller will use this bargaining power to extract some of the surplus generated by TP's high valuation. But this means that Seller and Buyer jointly enjoy a return on, say, an investment in Buyer's valuation even in the state of the world in which TP, not Buyer, acquires the good. So Buyer will overinvest.[81]

Do these arguments show that the penalty doctrine is efficient? They do show that, under certain conditions, enforcement of a liquidated damages clause can produce negative externalities. But the arguments do not travel the distance from this modest observation, to the conclusion that the penalty doctrine is justified. Indeed, the fit is poor. The penalty doctrine does not incorporate any of the variables identified in the literature: The cost of renegotiation, the distribution of valuations among potential TPs, the incentives to overinvest, and so forth. Further, the penalty doctrine effectively substitutes expectation damages for the invalid liquidated damages provision, but the literature we have been discussing does not establish that expectation damages are optimal.[82] Finally, parties can harm TPs in the way we have examined even without using liquidated damages, simply by overinvesting, which raises the expectation damages that the breacher would have to pay.[83] If courts care about efficiency and can detect this kind of strategic behavior, they should limit expectation damages. If they care about efficiency and cannot detect this kind of strategic behavior, they should not necessarily subject liquidated damages clauses to special scrutiny.

80. *See* Tai-Yeong Chung, *On the Social Optimality of Liquidated Damage Clauses: An Economic Analysis*, 8 J.L. ECON. & ORG. 280, 282-83 (1992).

81. *See* Kathryn E. Spier & Michael D. Whinston, *On the Efficiency of Privately Stipulated Damages for Breach of Contract: Entry Barriers, Reliance, and Renegotiation*, 26 RAND J. ECON. 180, 182-83 (1995).

82. *See* Chung, *supra* note 80, at 299.

83. *See* Tai-Yeong Chung, *Commitment Through Specific Investment in Contractual Relationships*, 31 CANADIAN J. ECON. 1057 (1998).

D. *Summary*

Our diversion through the literature on incomplete contracts has taken us through formidable terrain. This literature is flourishing, and I do not consider myself knowledgeable enough to criticize it.[84] Instead, I want to make a few points about its relevance for understanding contract law.

First, so far the literature has failed to predict the content of either contracts or legal doctrines such as the penalty doctrine.[85] Like the law-and-economics literature, the incomplete contracts literature founders on the ambiguity of contractual behavior and the difficulty of empirical investigation of this behavior. With no empirical basis for endorsing some assumptions and rejecting others, the models tend toward indeterminacy.

Second, if the literature has any normative implications, they are that courts should always specifically enforce all terms of every contract. Although we have seen highly stylized arguments that information asymmetries and externalities might provide reasons for judicial intervention, these arguments depend on implausible assumptions about the amount of information that is available to judges. Thus, we are left with a sterile normative defense of freedom of contract, one that is closely tied to its premises that parties know more about their interests than courts do.

Third, the problems with the literature suggest methodological complications for the theory of contract law. The literature takes more seriously than law and economics the premise that parties are rational, and permits them to design complicated contracts. But the premise of full rationality does not seem right, for it predicts contractual structures that bear little resemblance to the contracts designed by real parties. We will discuss these problems more fully in Section III.B.

84. But it should be noted that the literature is not free from controversy. Maskin and Tirole argue that if parties can commit not to renegotiate, the nonverifiability assumption does not explain the existence of incomplete contracts: Parties will use the same kind of contract regardless of whether the investment is verifiable. Maskin & Tirole, *supra* note 73, at 84. Hart and Moore reply by showing that if parties cannot commit not to renegotiate, nonverifiability does matter under certain conditions. Hart & Moore, *supra* note 73, at 116. The debate thus turns on whether parties can commit not to renegotiate, about which neither set of authors is able to marshal a decisive argument.

85. Or, for another example, courts are not likely to order parties to send messages (for example, name a price or pay a penalty) as required by some proposed contractual devices, such as that in Schwartz, *Incomplete Contracts*, *supra* note 24, at 281. We don't know this for sure, however, as these devices are not used, and thus are not the subject of judicial orders.

III. WHY ECONOMICS FAILS TO EXPLAIN CONTRACT LAW

A. *The Problem of Methodological Indeterminacy*

Richard Craswell has taken philosophical approaches to contract law to task for failing to provide fine-grained explanations of contract doctrine.[86] He points out that philosophical theories might explain in a general way why promises should be enforced—typically, by restating the moral intuition that promises should be kept, and then assigning the government a role in encouraging people to keep them—but never explain the details of doctrine. A theory that people should keep their promises does not tell us whether expectation damages, reliance damages, specific performance, or some other remedy is the appropriate response when a contract is broken. Indeed, when philosophers turn to these matters, they usually engage in implicit economic analysis or make assertions about the role of custom or other factors that are unrelated to their theories.[87]

Craswell's critique is methodological, not substantive. He argues that even if the philosophical theories capture some aspect of the truth about why contracts are enforced, they have no determinate implications for the phenomena that their authors purport to study—the doctrines, or the vast majority of the doctrines, of contract law. Although Craswell does not assert that economic analysis avoids this methodological problem, many readers will understand him to be implicitly making this claim. One cannot avoid being impressed by the contrast between the large and ingenious economic body of work on default rules, and the small and vapid body of work produced by philosophers.

But even if we accept Craswell's critique of philosophical theories of contract law, we must still ask whether economics really enjoys any advantages. By now, the answer should be familiar. Economists have proposed some models of contract behavior that make determinate but wrong predictions about the law. These models avoid Craswell's charge of indeterminacy, but they are still wrong. Determinate but wrong predictions enjoy a little more intellectual respectability than indeterminate predictions, but they get us not much closer to an understanding of contract law.

Economists have proposed other models of contract behavior that make predictions that are indeterminate. These models enjoy some intellectual advantages over the philosophical theories, for they would enable us to make complex and interesting predictions about contract law if we had

86. *See* Richard Craswell, *Contract Law, Default Rules, and the Philosophy of Promising*, 88 MICH. L. REV. 489 (1989).

87. *See, e.g.*, CHARLES FRIED, CONTRACT AS PROMISE: A THEORY OF CONTRACTUAL OBLIGATION (1981); Randy E. Barnett, *A Consent Theory of Contract*, 86 COLUM. L. REV. 269 (1986).

sufficient information about empirical conditions. But because we do not have such information, and it is—in my view, though others might disagree—unlikely that we ever would, the complex economic theories do not get us much closer to an understanding of contract law than the philosophical theories do.

Because this state of affairs could change with further research, one should not discount a renaissance in the law and economics of contracts. But another view is that theory and doctrine are mismatched, operating at different levels of generality. Welfare economics might be able to provide persuasive reasons for the superiority of a free market to, say, a planned economy. A free market can function only if people can trade, and trading almost always requires the making of binding promises. But there are many ways that promises can be made binding: through the operation of ordinary reputational mechanisms, through the creation of institutions like firms and trading associations that establish commitment mechanisms for members, and through contract law. And then there are many different rules of contract law that will be equally good at enabling people to make binding promises.[88] Specific performance is about the same as damages; literalistic interpretation is about the same as purposive interpretation. Individual contract doctrines, then, could be like rules of the road: sufficient as long as, within limits, everyone obeys them, and thus not susceptible to prediction on the basis of fine-grained theories of optimal interaction.

B. *The Problem of Rationality*

The economic scholarship on contract law purports to assume that individuals are rational in the sense of neoclassical economics. Their preferences obey certain consistency requirements, and their cognitive capacity is infinite. But on inspection, the nature of the rationality assumptions made by this scholarship is not so clear.

If individuals were rational, with no cognitive limits, and if transaction costs were zero, the role of contract law would be simple and uninteresting. Parties would foresee every possible future state of the world, and—the story goes—their contract would describe each party's obligation in each of these possible future states. For example, a contract for the sale of widgets would describe Seller's obligation if the cost of widgets increases or declines, and could make Seller's obligation turn on whether Seller invested in the right way, and so forth. Courts would specifically enforce the terms of the contract. In general conditions, efficiency would be obtained.

88. Richard A. Epstein, *Contracts Small and Contract Large: Contract Law Through the Lens of Laissez-Faire*, *in* THE FALL AND RISE OF FREEDOM OF CONTRACT 25, 26 (F.H. Buckley ed., 1999).

Economic analysis of contract law assumes that contracts cannot be designed to describe every future state of the world. The usual statement is that transaction costs prevent the parties from achieving such a detailed and complex contract. Some authors seem to mean the cost of negotiating and writing a contract; other authors seem to refer to cognitive limits of the parties, which include the inability to foresee future events and maybe something more.[89] In any event, one needs some such assumption to get the economic analysis of contract law off the ground; if the parties entered complete contracts, the law would not need to supply default terms such as expectation damages. Instead, the parties would choose expectation damages whenever they anticipated the need, namely, when they wanted to give the promisor the option to perform or pay a sum of money to the promisee.

Let us examine the two main ways that authors use the idea of transaction costs. The first approach assumes that parties are rational but that entering a contract involves some special cost. Some authors assume that this cost refers to time spent negotiating or the time and materials needed to draft a document.[90] Others assume that the cost results from problems of asymmetric information, and, in particular, the inability of a court to verify a subset of the contract-related actions in which the parties engage.[91] As discussed in Part II, rational parties would minimize these costs by entering contracts that incorporate complex ex post bargaining mechanisms. Yet there is little evidence that such mechanisms are used in real contracts.

This problem is less clear in the law-and-economics literature than in the incomplete contracts literature, but that is only because the law-and-economics models constrain the types of contracts that parties may enter rather than formally modeling the transaction cost. The models generally permit parties to choose prices and quantities in noncontingent contracts, and not to choose contracts that stipulate ex post bargaining procedures, though these are likely to be superior. One defense of the methodological approach of law and economics is that the latter contracts are too complex to be useful, so it is a justified simplification to assume that only fixed price contracts are available to parties.

This brings us to the second approach to the idea of transaction costs, which is, in fact, to treat it as a metaphor for bounded rationality, the idea that parties are rational to the extent permitted by limits on cognitive capacity.[92] Although law-and-economics scholars rarely put their argument in this way, the assumption is reflected, as just noted, in the modeling

89. *See, e.g.*, COOTER & ULEN, *supra* note 2, at 100, 236, 240.
90. *E.g.*, Ayres & Gertner, *Filling Gaps*, *supra* note 16, at 93.
91. Like investment levels. *See* Hart & Moore, *supra* note 61.
92. *See, e.g.*, WILLIAMSON, *supra* note 61, at 21.

device of permitting parties to choose only noncontingent contracts.[93] To understand the problem with this strategy, consider the common claim that default rules should be designed to give a party an incentive to reveal information about the party's cost or valuation. If "transaction costs," meaning bounded rationality, prevent the parties from choosing a sophisticated contract in light of future events, then they should also prevent parties from anticipating the effect of legal rules (which would be applied only in the contingent future) on the simple contract that they design. Instead, the model simultaneously assumes that individuals can foresee remote events and make complex calculations (otherwise they would not be motivated by the default rule to release information) and cannot engage in a perfect cognitive response (otherwise the cost of entering the bargain would be zero). The assumptions are jointly implausible.[94]

Economists reject bounded rationality arguments for two reasons. The first is methodological: They cannot agree on a standard, mathematically tractable formulation of bounded rationality. This might be a good reason for economists, but it is a bad reason for lawyers. The second is empirical: If the rationality assumptions of economics are close enough to the reasoning of individuals, or to the institutionalized reasoning implemented by firms, then the conclusions of the economic models are also good enough for predictive and normative purposes. My view is that the failure of contracts to include the mechanisms identified by the incomplete contracts literature is evidence that the rationality assumptions are not good enough. Others might disagree, claiming either that transaction costs—that is, high writing costs, severe information asymmetries, etc.—explain the absence of these mechanisms, or that better modeling will lead to different conclusions in the future. This seems to me a dodge, especially in the absence of an empirical test of the role of transaction costs in preventing the use of mechanisms.[95] But it cannot be dismissed out of hand. The question, then, is whether one should have optimism or pessimism about future research.

93. I do not mean that the authors self-consciously made this modeling choice as a way of capturing bounded rationality. I mean that they usually do not explicitly model transaction costs as, for example, the cost of drafting the contract and, instead, treat transaction costs as an (informal) explanation for why they assume that parties can choose only the contract price and quantity (and sometimes a liquidated damages provision). *E.g.*, Shavell, *supra* note 11. Even when a variable is used to refer to a transaction or communication cost in a formal model, the parties are constrained in their choice of contractual form. *E.g.*, Ayres & Gertner, *Filling Gaps*, *supra* note 16, at 108.

94. A similar point has been made about the incomplete contracts literature. *See* Maskin & Tirole, *supra* note 73, at 84; Segal, *supra* note 64, at 74.

95. But, of course, if there are no contracts that use these mechanisms, the techniques of econometrics are hardly necessary.

C. *A Way Out?*

One way out of this impasse involves greater consideration as to what parties can realistically be expected to foresee. Contrast the efficient breach theory and the *Hadley* theory. If Seller experiences higher than expected costs and would like to avoid the deal, she would probably consult a lawyer. The lawyer would tell her that if she breaches the contract, she will probably pay expectation damages. Comparing the cost of performance and the expectation damages, Seller will decide whether or not to perform. This decision seems well within the cognitive abilities of an ordinary Seller.

By contrast, the *Hadley* theory applies to a decisionmaking process that occurs well before performance. At the time of contracting, Buyer (for example) must anticipate that Seller might breach rather than perform; that Seller must pay damages if she breaches; that these damages depend on the valuations of other buyers as well as Buyer's revelation of his private valuation; that Seller will (or maybe won't) anticipate these damages when deciding how much care to use when performing; that Seller will (or maybe won't) use Buyer's information to price discriminate; and so on. This chain of reasoning seems likely to exceed the cognitive capacities of an ordinary Buyer, whether or not a lawyer is consulted.

One way out of the impasse, then, requires incorporation of cognitive limitations into a theory of the relationship between contract law and contract-related behavior, so that one can distinguish incentives that are likely to influence behavior, and those that are too remote to influence behavior. No widely accepted theory of bounded rationality exists, however. The likelihood that such a theory could be developed is discussed in Section V.D.

IV. THE INFLUENCE OF ECONOMICS ON CONTRACTS SCHOLARSHIP AND LAW

A. *The Influence of Economics on Contracts Scholarship*

Defenders of economic analysis of contract law point to the significant influence of economics on contracts scholarship. Indeed, this influence can be documented in many ways. Economic analysis played almost no role in contracts scholarship prior to 1970, whereas it has played a dominant role in much of the contracts scholarship published in the major law reviews in the 1990s. It has influenced some of the analysis in the treatises. It shows up in the casebooks. Economic articles on contracts are frequently cited in noneconomics articles. And contracts scholars at the top law schools are frequently identified with the law-and-economics approach.

But this rosy picture does not tell the whole story. The most influential economic articles, with one exception, were published in the 1970s and early 1980s. The exception, Ayres and Gertner's 1989 article on default rules, is usually cited for the useful distinction between majoritarian and penalty defaults and not for the economic analysis of this distinction. On the whole, economics-of-contracts articles published in the last fifteen years are cited no more frequently than noneconomics articles are.[96]

Contract-law casebooks and treatises show the influence of economics, but it is the influence of pre-1980 economics. Most casebooks and treatises mention the idea of efficient breach, but not the equally important idea of efficient reliance.[97] Casebooks generally treat the economic approach as an exotic "perspective," as an object at which to marvel, and not as the underlying logic of contract law. To be sure, most casebook authors are noneconomists, but what is important is that these authors have apparently concluded that greater economic content would not expand the market share of their casebooks.[98]

B. *The Influence of Economics on Contract Doctrine*

The influence of economic analysis on contract law is harder to discern. Let us start with the common law. Judicial opinions occasionally cite economic articles, and occasionally use economic concepts such as transaction costs and risk aversion. But it is hard to find cases where the judges self-consciously rely on an economic argument in order to justify a result. One such case is the *Van Wagner* case,[99] which relies heavily on Anthony Kronman's analysis of specific performance.[100] Many opinions

96. This conclusion is based on a regression of the annual citations of articles published since 1980 in leading law journals (California, Chicago, Columbia, Harvard, Michigan, Northwestern, NYU, Pennsylvania, Stanford, Texas, UCLA, and Yale) and faculty-edited journals (*Journal of Law and Economics*, *Journal of Law, Economics, and Organization*, and *Journal of Legal Studies*) on a dummy variable equal to one if the article uses economic analysis, and zero otherwise. The mean annual citation for economics articles was 3.8 (71 articles); for noneconomics articles, it was 4.1 (52 articles). If you exclude all the faculty-edited journals, however, then economics articles are cited more often at a statistically significant level; but if you include the *Journal of Legal Studies* alone, then they are not. There are many ways that one could conduct this test, and for that reason I can conclude only that citation evidence does not exclude the hypothesis that economics articles are no more influential than noneconomics articles.

97. A Westlaw search of Farnsworth's treatise yields the following results: "efficien!" = 32 results; "economic!" = 113 results; "transaction! cost!" = 5 results; "Coase" = 2 results. The "economic!" search caught many concepts unrelated to economic analysis, such as "economic waste."

98. The Scott and Leslie casebook is the exception; compare the contracts casebooks by Barnett, Dawson, and Eisenberg. *Compare* ROBERT E. SCOTT & DOUGLAS L. LESLIE, CONTRACT LAW AND THEORY (2d ed. 1993), *with* RANDY E. BARNETT, CONTRACTS: CASES AND DOCTRINE (1995), JOHN P. DAWSON ET AL., CONTRACTS: CASES AND COMMENT (7th ed. 1998), *and* LON L. FULLER & MELVIN ARON EISENBERG, BASIC CONTRACT LAW (6th ed. 1996).

99. Van Wagner Adver. Corp. v. S & M Enters., 492 N.E.2d 756, 760 (N.Y. 1986).

100. Kronman, *supra* note 14.

cite Posner and Rosenfield's article on the impossibility doctrine, but almost always for the proposition that contracts shift risk, an idea that predates economic analysis by many decades. The notes to the *Restatement* contain only a handful of references to economic ideas.[101]

To examine the influence of economic analysis on contract doctrine more systematically, I read the state and federal court opinions that cite an economics article that appeared in a major law review or faculty-edited law journal after 1980. Only thirty-six such opinions were issued. Of these, few discussed rather than cited the article, and none was clearly influenced by an article.[102]

Economic analysis has also had little influence, as far as I can tell, on statutory and regulatory law. Statutes and regulations in the 1970s incorporated common-law developments that economic analysis criticized, and although the consumer-protection movement crested in the 1970s, economics did not spur deregulation of the consumer-product and consumer-finance markets as it did for so many other markets such as trucking and air travel.

V. THE FUTURE

If the limits of economic analysis are becoming visible, can some other methodology take us beyond them? A brief survey of the other contenders provides grounds for concern.[103]

A. *Philosophy*

For a long time, legal scholars have sought a philosophical explanation for contract law. Fuller and Perdue argued that contract law is based on corrective justice.[104] Fried argued that contract law is based on the morality

101. RESTATEMENT, *supra* note 48. Searches on Westlaw yielded the following results: "efficien!" = 8 results; "economic!" = 38 results; "transaction! cost!" = 1; "Coase" = 0 results. In fact, few of the "economic!" results were related to economic analysis (as opposed to "economic waste," etc.).

102. Richard Posner's *Economic Analysis of Law*, however, has been cited in 116 cases in the LEXIS contracts database. A sample suggests a mix of meaningful analysis and meaningless citation. Cooter and Ulen's textbook was cited only twice. It might be the case that economically minded judges such as Calabresi, Easterbrook, and Posner have influenced contract doctrine, and I have not tried to measure their influence by looking at whether other judges accept their views about contract law. William Dodge credits the theory of efficient breach as provoking a judicial backlash against awarding punitive damages for breach of contract. *See* William S. Dodge, *The Case for Punitive Damages in Contracts*, 48 DUKE L.J. 629, 642-44 (1999).

103. For a more optimistic view, see ROBERT A. HILLMAN, THE RICHNESS OF CONTRACT LAW (1997).

104. Lon L. Fuller & William R. Perdue, Jr., *The Reliance Interest in Contract Damages*, 46 YALE L.J. 52 (1936).

of promising.[105] These theories remain the most influential despite their inadequacies.[106]

Let us first consider each theory from a descriptive perspective. Fried argues that contracts should be enforced because individuals have a moral obligation to keep their promises. Fried's theory has the virtue of simplicity but cannot explain the many ways that contract law refuses to enforce promises. Unreciprocated offers, promises that lack consideration, promises that violate the Statute of Frauds, promises that lack specific terms—all of these promises are, in ordinary cases, not enforced. Finally, as Craswell has pointed out, Fried's theory cannot explain the default terms the law uses to fill out promises that otherwise are ill-defined.[107]

Fuller and Perdue argue that contracts should be enforced in order to prevent one party, the promisor, from benefiting at the expense of the other party. Corrective justice demands that the breaching promisor make the promisee whole. The reliance measure is ideal for this purpose, but because reliance costs are hard to measure and the expectation measure approximates the reliance measure in competitive markets, the expectation measure is the appropriate rule. But as Craswell shows, Fuller and Perdue's theory cannot explain why the appropriate baseline for exercising corrective justice is the promisee's position prior to the making of the promise, as opposed to after the making of the promise.[108] Others—notably Grant Gilmore and Patrick Atiyah—tried to generalize Fuller and Perdue's analysis and claimed that contract is being absorbed into tort.[109] But these efforts have gone nowhere.

The theories fare no better when conceived as normative arguments for the reform of contract law. As Craswell shows, they are indeterminate over nearly all aspects of contract doctrine.[110] Fried's theory justifies the enforcement of promises, but sheds no light on which of many remedies—expectation damages, reliance damages, specific performance, even nominal damages—is the right one. Fuller and Perdue's theory, as just mentioned, cannot solve the baseline problem.

105. FRIED, *supra* note 3.

106. The present discussion concerns normative theories, not "analytic" or "interpretive" theories (as Craswell uses the terms) that have been advanced by legal philosophers. For the distinction, see Stephen Smith, An Introduction to Contract Theory (2002) (unpublished manuscript, on file with author). Philosophers have goals that are different from those of economists, *see* Jody S. Kraus, *Philosophy of Contract Law*, *in* THE OXFORD HANDBOOK OF JURISPRUDENCE AND PHILOSOPHY OF LAW 687 (Jules Coleman & Scott Shapiro eds., 2002), and here I am asking only whether philosophy can succeed where economics has failed.

107. Craswell, *supra* note 86, at 521-23.

108. *Id.*

109. P.S. ATIYAH, ESSAYS ON CONTRACT (1990); GRANT GILMORE, THE DEATH OF CONTRACT (Ronald K.L. Collins ed., 1995).

110. Craswell, *supra* note 86.

None of this is to say that philosophy has nothing to offer contract law. Philosophical reasoning, if not necessarily the reasoning of philosophers, has a significant accomplishment: the critique of the "will theory" of contract. The will theory, which derives contract doctrine from the premise that a contract is the coming together of two wills, is not just a once popular legal theory; it is also an intuitive, common sense approach to understanding contract law, instinctively adopted by generation after generation of first-year law students, and a happy target for philosophical criticism. The celebrated critique of the will theory of the duress doctrine—namely, contracts entered under duress and contracts entered "voluntarily" involve the same kind of coercion—is powerful and important, and this critique owes something to philosophical reasoning. But one must also understand that this critique is much older than those who are credited with it: It can be found not only in Dawson, and Hale, and Holmes, but also in Hume and probably earlier.[111] Basic philosophical ideas about the nature of the will, of agreement, and of contract go back for centuries, and their implications for contract law are well understood.

B. *Psychology*

Several scholars have recently argued that cognitive psychology holds promise for explaining the law, including the law of contracts. This view has superficial attractiveness.[112] If, as I have argued, economic models of the law are undermined by their rationality assumptions, then psychologically accurate models of human cognition might fill in the gaps left by the economic explanation.

Let us focus on the example of the penalty doctrine because it is the topic of a recent debate about the value of using cognitive psychology to understand the law. We have already seen that economics fails to explain the penalty doctrine; can cognitive psychology?

The question is, why do courts give less deference to liquidated damages clauses than they do to other provisions of a contract, including choice-of-forum clauses, which will become relevant only if a dispute arises? In response to Robert Hillman's skepticism about whether cognitive psychology can explain this practice,[113] Jeffrey Rachlinski argues that (1) biases that cause overoptimism justify scrutiny of liquidated damages provisions; (2) the status quo bias (contrary to Hillman's claim) does not

111. DAVID HUME, A TREATISE OF HUMAN NATURE 525 (L.A. Selby-Bigge ed., Oxford Univ. Press 1978) (1740).

112. *See, e.g.*, Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CAL. L. REV. 1051 (2000).

113. Robert A. Hillman, *The Limits of Behavioral Decision Theory in Legal Analysis: The Case of Liquidated Damages*, 85 CORNELL L. REV. 717 (2000).

justify deference because the increased effort to bargain around the damages rule does not necessarily eliminate the effects of overoptimism; and (3) although aversion to ambiguity justifies deference to liquidated damages, courts actually use this insight under the penalty doctrine by giving more deference to liquidated damages clauses when damages are hard to calculate (and thus ambiguous).[114]

Even accepting these arguments, which will strike many as ad hoc, Rachlinski cannot explain why the biases justify judicial scrutiny of liquidated damages terms but not other terms. Breach is not the only low-probability event that occurs within contractual relations; a contract might make any number of obligations conditional on events that occur with a low probability. Think of bond covenants that give creditors the right to accelerate repayment when the debtor's asset-debt ratio falls below a threshold, employment compensation packages that provide payoffs only when market conditions are favorable, and sales contracts that allocate the risk of the destruction of the goods during delivery. If parties overlook low-probability events, then any of these provisions could be defective, but because they are not liquidated damages provisions, courts do not subject them to scrutiny. Indeed, Rachlinski concedes the explanatory failure of cognitive psychology when he says that the field "might cause scholars to question much of contract law's foundations."[115] Rachlinski slips from a descriptive claim to a normative claim in the face of the poor fit of cognitive psychology and the penalty doctrine.

C. *History*

Historical explanations of contract law once held promise, but early enthusiasm has given way to skepticism. Consider the attempt to link trends in contract-law doctrine to the rise of the welfare state.[116] Scholars claimed that the increasing informality of contract law over the last century, and especially the rise of promissory estoppel, showed courts moving away from laissez faire and toward statism and the enforcement of community standards. The convenient link to other trends in political economy, and specifically the rise of the welfare state, obscured the poor fit between the theory and doctrinal trends.[117] The rise of promissory estoppel, for example, could be interpreted as reflecting judicial impatience with a formality—the consideration doctrine—that interfered with, rather than promoted, private

114. Jeffrey J. Rachlinski, *The "New" Law and Psychology: A Reply to Critics, Skeptics, and Cautious Supporters*, 85 CORNELL L. REV. 739, 762 (2000).

115. *Id.* at 763.

116. P.S. ATIYAH, THE RISE AND FALL OF FREEDOM OF CONTRACT (1979); GILMORE, *supra* note 109.

117. *See* Epstein, *supra* note 88, at 25.

contracting.[118] Contract doctrine can coexist with many different political systems, and broad trends, such as the decline of formalism, do not necessarily reflect changes in politics or morality.

Similar criticisms apply to Simpson's argument that the eighteenth-century shift from judicial accommodation of penalties to hostility toward penalties was due to "social evolution" away from tolerance for the private use of terror to the monopoly of force held by the state.[119] Simpson argues that once courts deprived private parties of the right to use force, the courts vindicated their longstanding commitment to the compensation principle by banning penalties.

The argument raises more questions than it answers. Both before and after the "social evolution," parties depended on courts to enforce their contracts. Before, a party could not collect a penal bond without first obtaining a judgment from a court. The other party had a number of legal defenses: not only full performance of the underlying promise, but such conventional defenses as duress and impossibility. The doctrinal change did not reflect a shift away from tolerance for private use of terror; it reflected a shift in the degree of deference given to remedial terms in contracts. Simpson's argument boils down to the assertion that judges stopped deferring to remedial terms because they wanted to control remedies, but he does not explain why they would want to treat contractually stipulated remedies differently from other terms in the contract.

Even if we accepted the "social evolution" argument—the shift from private to public remedies—we need to understand why judges would think that the "compensation principle" should control remedies and thus exclude penalties.[120] Simpson argues that judges had a longstanding belief that the law should provide compensation (not overcompensation) for injuries. But judges also believed that "the real function of contractual institutions is to make sure, so far as possible, that agreements are performed."[121] Simpson acknowledges that the two principles—compensation and respect for agreement—are in conflict, but does not explain why the first prevailed over the second.

Simpson's argument shares the flaws of the historical scholarship described above—the use of macro trends to explain micro phenomena that are consistent with other trends, the casual appeal to such long-term trends to explain a change that occurred at a particular time, and the arbitrary resolution of tensions between different principles or ideas in favor of one

118. Eric A. Posner, *The Decline of Formality in Contract Law*, *in* THE FALL AND RISE OF FREEDOM OF CONTRACT, *supra* note 88, at 61.

119. A.W.B. SIMPSON, A HISTORY OF THE COMMON LAW OF CONTRACT 124 (1975).

120. Recall also the ambiguity of the notion of "compensation," which presupposed a baseline. *See supra* text accompanying note 108. But we will assume that whatever compensation means, a penalty requires something beyond compensation.

121. SIMPSON, *supra* note 119, at 123.

rather than the other. Historical scholarship is often illuminating, and any good theory of contract law would need to account for aspects of its historical evolution, but thus far theories of contract law emerging from historical research have not resolved basic puzzles about modern contract doctrine.

D. *Advances in Economics and Game Theory*

I argued in Section III.C that the failures of the economic analysis of contract law derive, in part, from the bounded rationality of individuals who enter contracts. If people were rational, then their contracts could be predicted. And if people's behavior could be predicted in this way, then firm recommendations about contract doctrine could be made. But because the models press rationality to its limits, the world falls short of the predictions, and so the natural consequence is to incorporate cognitive limitations into models of behavior.

The economic literature on bounded rationality is complex and large, and I cannot do justice to it. None of the models of bounded rationality that have been proposed has achieved canonical status,[122] and thus it is difficult to discuss in general what bounded rationality means for contract law. An example will illustrate the problems, and, I think, justify skepticism about the ability of future models of bounded rationality to shed light on contract doctrine.[123]

In the beauty contest game, the experimenter asks each member of a group to write down a number between 0 and 100 that is, say, 2/3 of the average number (between 0 and 100) that everyone else writes down. The person who writes the correct answer wins a prize; if there is a tie, the prize is divided among the people with the right answer.

Game theory—that is, game theory that assumes "unbounded" rationality—predicts that everyone will write down 0 and share the prize. The intuitive explanation for this prediction follows. Imagine that you are one of the people asked to write down the number. You might start by imagining that everyone else will pick a number at random. If so, you might expect a uniform distribution from 0 to 100, with a mean of 50. Thus, you would pick 2/3 of the mean, which is about 34. But then it might occur to you that everyone will have reasoned in the same way that you have. Thus, everyone else will have written down 34. But if everyone else has written down 34, then you can win only by writing down 2/3 of 34, which is about 23. Yet everyone else knows this as well, so they will write down 23, and

122. *E.g.*, David M. Kreps, *Bounded Rationality*, *in* 1 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW, *supra* note 5, at 168.

123. For other efforts, see *id.* at 168-73.

you should write down 2/3 of 23, and so on. Continuing with this reasoning, we reach 0. If everyone chooses 0, then you can do no better than choosing 0 as well, for then you will share the prize, whereas if you choose a higher number, you will receive nothing.

The explanation can be given more formally. If everyone chooses 0, everyone gets a share of the prize. If one person decides to choose a number different from 0, that person will no longer receive a share of the prize, and thus will receive a lower payoff. No one can do better than choosing 0 given that everyone else chooses 0. The outcome 0 is what game theorists call a Nash equilibrium: It is an outcome from which one has no incentive to deviate, given everyone else's choice, because one cannot increase one's payoff by deviation.[124] By contrast, if one person knows that everyone else will choose a particular number $n > 0$, then that person can do better by choosing a number different from n, namely 2/3 of n. Thus, any $n > 0$ cannot be a Nash equilibrium.

Nash equilibrium does a poor job predicting behavior in the initial rounds of play. When people play the beauty contest game, typically there is a distribution as follows. For a sufficiently large group, most or all numbers will be chosen, with spikes around 34, 23, and perhaps 14, and then 0.[125] A natural explanation of this pattern is that some people choose numbers randomly, or misunderstand the game; others are able to think one, two, or three steps ahead, or even more, but it never happens that everyone figures out, and plays, the Nash equilibrium in the initial rounds.

Aside from the empirical disconfirmation, the experiments pose a conundrum. Suppose we imagine a perfectly rational person, X. What do you predict that X will do? In Nash equilibrium X chooses 0, but if X is smart (rational?) enough, X will realize that not everyone else will play the Nash move, in which case X should choose a number greater than 0, its magnitude depending on the distribution of cognitive ability in the population. So our purely rational X will not act purely rationally as defined by game theory.

What to do? One idea recently investigated by Teck-Hua Ho and his coauthors proceeds as follows.[126] Imagine that a person engages in a number of cognitive steps when thinking about how to play the beauty contest game. In step 0, he randomly chooses a number between 0 and 100. In step 1, he thinks that everyone else has engaged in step 0 and only step 0—that is, everyone else has chosen a number at random—and chooses

124. *See* DOUGLAS G. BAIRD ET AL., GAME THEORY AND THE LAW 19-23 (1994).

125. *See* Teck-Hua Ho et al., *Iterated Dominance and Iterated Best Response in Experimental "p-Beauty Contests,"* 88 AM. ECON. REV. 947, 953-58 (1998). Their version of the game uses a different range of numbers from the example in the text, which is simpler, but the pattern is the same.

126. *See id.*

strategically on the basis of this assumption. (In our example, he would choose 34.) In step 2, he thinks that a fraction of the population has engaged in step 0, and the rest has reached step 1, but none (besides himself) has reached step 2, and he chooses accordingly. This process continues for an arbitrary number of steps.

To predict the distribution of numbers chosen in the beauty contest game, we assume that a portion of the population stops at step 0, another group stops at step 1, another group stops at step 2, and so forth. To make such a prediction, we have to decide how many steps there are, and how the cognitive ability (the number of steps taken) is distributed among the population, but let us suppose that we can make reasonable assumptions about these parameters. Ho and his coauthors show that if we make such assumptions, we will predict something close to the actual distribution—all or most numbers being chosen, with spikes near 34, 23, and so forth, except that there will not be a spike at 0. (If the game is repeated, however, people will learn from their mistakes and eventually nearly everyone will play the Nash equilibrium strategy.)

The model has some attractive features: It captures the importance of the distribution of cognitive capacities that presumably exists in the general population, the effect of limited cognitive capacity on choices, and the role of learning. All of these factors must play a role in the design of contracts and therefore in the proper judicial treatment of them. But the model does not refine the basic intuition with which we started, that contract doctrine might have something to do with mistake, lack of foresight, and similar effects of cognitive limitations.

The role of a model of bounded rationality in normative analysis of contract law is also obscure. If parties cannot foresee certain events, then legal rules will not affect their incentives, and courts can do what they want when those events occur. If parties can foresee the events but fail to think about them fully and accurately, then the possibility of useful judicial intervention remains open. But an accounting of the costs and benefits of this intervention must await a more fully worked out theory of bounded rationality.

E. *A Return to Doctrinalism?*

If interdisciplinary approaches to contract law cannot generate plausible descriptive or normative theories, should legal scholars return to doctrinal analysis? To answer this question, we must first be clear about what doctrinalism means.

The most ambitious doctrinal scholarship attempts to derive principles from cases. Fuller and Perdue argued that a "reliance principle" explains contract damages.[127] More recently, Eisenberg has argued that a "bargain principle" and a "fairness principle" explain contract-enforcement doctrines,[128] and Farnsworth has proposed "dependence" and "public interest" principles, among many others.[129]

As many have observed, cases will not yield principles that are more general than the case outcomes themselves. The plausibility of the principles that scholars advance always comes from their appeal to moral commitments. The extraction of a fairness principle from the cases, rather than a principle of fairness for litigants who have brown eyes (if such is the case) or for litigants who have some other characteristic coextensive with the cases in which they prevail, is always the result of an implicit appeal to an attractive normative idea—fairness for all, rather than fairness to a morally arbitrary group of people. Ambitious doctrinal scholarship thus converges to a kind of moral philosophy that is especially sensitive to judicial outcomes,[130] and is thus vulnerable to the criticisms of philosophical analysis described in Section V.A.

Examples of authors who write in this vein include Fuller and Perdue, who argue that the reliance principle follows from corrective justice, and Eisenberg, who attempts to tie the bargain principle and fairness principle to policy concerns and moral commitments. Farnsworth avoids philosophizing or engaging in policy analysis by keeping his discussion vague. For example, he does not explain how he resolves conflicts between the many principles that he invokes.[131]

A narrower kind of doctrinal analysis is nothing more than ordinary legal analysis, in which a judge or lawyer explains whether or not a given precedent controls the case under consideration. This kind of doctrinalism is useful, and can be done well or poorly, but a return to this scholarship would have to count as a defeat for the descriptive and normative aspirations of modern legal theory. Doctrinalism does not describe or justify the law; it is simply the use of legal materials and techniques of reasoning to determine the outcome in a given case, or to reconcile or criticize cases that have disparate outcomes.

---

127. Fuller & Perdue, *supra* note 104.

128. Melvin A. Eisenberg, *The Bargain Principle and Its Limits*, 95 HARV. L. REV. 741 (1982).

129. E. ALLAN FARNSWORTH, CHANGING YOUR MIND: THE LAW OF REGRETTED DECISIONS (1998).

130. That is, something like Dworkinian interpretivism.

131. For a critique of Fuller and Perdue's argument, see Richard Craswell, *Against Fuller and Perdue*, 67 U. CHI. L. REV. 99 (2000). For a critique of Farnsworth's argument, see Eric A. Posner, *Law and Regret*, 98 MICH. L. REV. 1468 (2000).

VI. CONCLUSION

The title of this Essay is a question, not a statement, for two reasons. First, the answer can come only with more experience. As economists and lawyers experiment with new models and variations of old ones, they might find better approaches to understanding contracts and contract law. The answer might also turn out to be "sort of," depending on whether efforts to model bounded rationality, should they succeed, ought to be considered a vindication of economics or psychology.

Second, economics has already accomplished much, just not what its proponents set out as the measure of success. If you look at the best work in contracts scholarship today and compare it with the best contracts scholarship before the 1970s, you will see many differences. One important difference is that earlier work was methodologically sloppy. Much of this work mixed up two separate tasks: excavating the doctrine and evaluating it. Evaluation would often be based on poorly articulated notions of fairness—intuitions that other commentators as well as judges might or might not share—with either no attention to the effects of doctrine on incentives, or casual discussion. The failure to distinguish doctrine and policy often resulted in the displacement of the policy disagreement into rule/standard debates. For example, the voluminous literature on the unconscionability doctrine in the 1960s and 1970s was vague on the incentive effects of the doctrine—with some concern about interfering with freedom of contract, and some concern about unequal bargaining power—and vigorous on the question whether an ambiguous standard like unconscionability could be applied by courts as consistently as they applied similar supposedly rule-like doctrines, such as duress.[132] No one seemed to understand that the rules/standard question presupposed a resolution of the policy question.

The economic literature on the unconscionability doctrine clarified the policy questions at stake and largely displaced the earlier literature. Its main accomplishment was showing that the earlier policy arguments were ill-defined, or made implausible empirical assumptions, or were inconsistent with widely held views or other uncontroversial areas of law and policy. Defenses of the unconscionability doctrine are now more candid and clearer, even if they reject economic premises. The literature as a whole proceeds at a higher level of sophistication.

Economics, then, ushered in a set of scholarly virtues that have improved the literature. These virtues include consistency in the use of terms, clarification of the stakes of the discussion, the distinction between normative and positive analysis, and isolation of different incentives and

132. *See, e.g.*, M.P. Ellinghaus, *In Defense of Unconscionability*, 78 YALE L.J. 757 (1969).

behaviors. The literature now speaks in an economic idiom, with concepts like transaction cost, risk aversion, default rule, and efficiency substituting for similar but vaguer notions in the earlier writings. These are important accomplishments, and it is hard to imagine serious contract-law research in the future that does not reflect the influence of economics.

But economics fails to explain contract law. It does not explain why expectation damages are the standard remedy, for example, or why liquidated damages are not always enforced. It does not explain the function of the consideration doctrine or promissory estoppel. It does not explain why the law sometimes encourages people to disclose information and at other times does not.

And economics provides little normative guidance for reforming contract law. Models that have been proposed in the literature either focus on small aspects of contractual behavior or make optimal doctrine a function of variables that cannot realistically be observed, measured, or estimated. The models do give a sense of the factors that are at stake when the decisionmaker formulates doctrine, and might give that decisionmaker a sense of the trade-offs involved, but in the absence of information about the magnitudes of these trade-offs—and the literature gives no sense of these magnitudes—the decisionmaker is left with little guidance.