What does an artificial design agent mean by being 'situated'?

Gregory J. Smith and *John S. Gero*, Key Centre of Design Computing and Cognition, University of Sydney, NSW 2006, Australia

Schön described designing as a 'conversation with materials conducted in the medium of drawing'. Both the problem and solution of many designing tasks emerge through this 'conversation' between a situated designer and the medium of the design. Unfortunately, describing agents as 'situated' means different things to researchers from different fields. In this paper we review work from different fields so as to describe what 'situated' means for a design agent.

Keywords: situated agent, computational models, design theory

Designing is the conscious effort to impose meaningful order (Margolin, 1986, quoting Victor Papenek). Conceptual designing is an early phase of design which is characterised by abstractness and an incomplete understanding of the problem and/or solution (Gero, 1998). Designers cope with this by exploring the space of design requirements at the same time as they begin to try and understand the space of conceptual designs. This is achieved by interacting with the media of the conceptual designs as exemplified by Schön and Wiggins' (1992) 'conversation with the medium'. Designers, human and artificial, have therefore been described as situated agents. But what is meant by 'situated' varies across disciplines. What Suchman (1987) understands by the term 'situated' has a cognitive and sociological character that is distinctly different from a common AI understanding that equates it with 'embodied'. So, when talking of design agents, what does 'situated' mean?

The traditional computational approach presumes that designing is search and planning; achieving design goals through internal reasoning with inference rules over models in a suitable logic or language. These methods search an encoded space for a goal state, and require good heuristics to be effective. There is no notion of interaction revealing

Corresponding author:
G. J. Smith
g_smith@arch.usyd.edu.au

www.elsevier.com/locate/destud
0142-694X $ - see front matter *Design Studies* **26** (2005) 535–561
doi:10.1016/j.destud.2005.01.001

alternatives not encoded in the space. It is an approach characterised by Coyne et al. (1990).

A logical proposition, however, is not necessarily the same as a design proposal as design problems often cannot be comprehensively stated (Lawson, 1997). Some aspects of a particular design problem do not emerge until an attempt has been made to solve it. So how does search or planning in a solution or plan space account for designing if knowledge of the design environment or design goals are incomplete or in error, or if design knowledge itself contains errors/omissions, or is incomplete, or if the design task can be formally stated but planning is exponentially complex and over a large solution space?

1 Interaction

The heuristics that are required of any search are a part of the domain and common-sense knowledge available to a designer. Expert systems programmers have been trying to program disembodied common-sense knowledge for decades (Horgan, 2004). One reason suggested by Brooks (1995), Clancey (1997) and others for why this has been a struggle is precisely because their systems are not situated and embodied. By contrast with conventional planning, in a situated view plans are constructed as an artifact of '*reasoning about* action, not the generative *mechanism of* action' (Suchman, 1987, emphasis is Suchman's).

An example is of a designer and sketches made during the early conceptual phases of a design task. There is a difference between this viewed as an interaction with a drawing and, say, viewing it as searching of encoded model of a drawing. One difference is that expectations of what is in a drawing influence how it is perceived, and this influence feeds back into ongoing perceptions of that drawing. Consider the scene shown in Figure 1. We naturally believe that the man is contemplating jumping from the ledge. Now look at Figure 2, which is the same scene a few seconds later. The reason that this is funny is that it contradicts our expectations.

Consider now Figure 3(a). We do not simply look at Figure 3(a) and parse what is sensed into objects; we interact with the figure. Biasing our perception are expectations of what will be perceived. The concepts that Dali had in mind when he produced Figure 3(a) most likely include concepts of a greyhound, the mythological beast and so on as well as others that associate in his mind with those.

Figure 1 Man standing on window ledge, from Milligan and Shand (1996)

What we conceive of while interacting with this figure, as a viewer and independent agent, is not necessarily as Dali intended. Indeed, it is not assured that what the painter conceives of afterward is only that originally intended. Many people viewing Figure 3(a) for the first time will not find all of the interpretations intended by Dali without the assistance of Figure 3(b) sketches. Equally, just because Dali produced the image does not mean that those six interpretations are exhaustive or even necessarily correct. How we interpret the figure depends on our expectations, the current situation, and we construct the memories, beliefs and expectations that bias our perception.

Our work has therefore been motivated by a desire for a model of designing that is based on interaction; of a situated agent that can

Figure 2 The scene of Figure 1 but a few seconds later

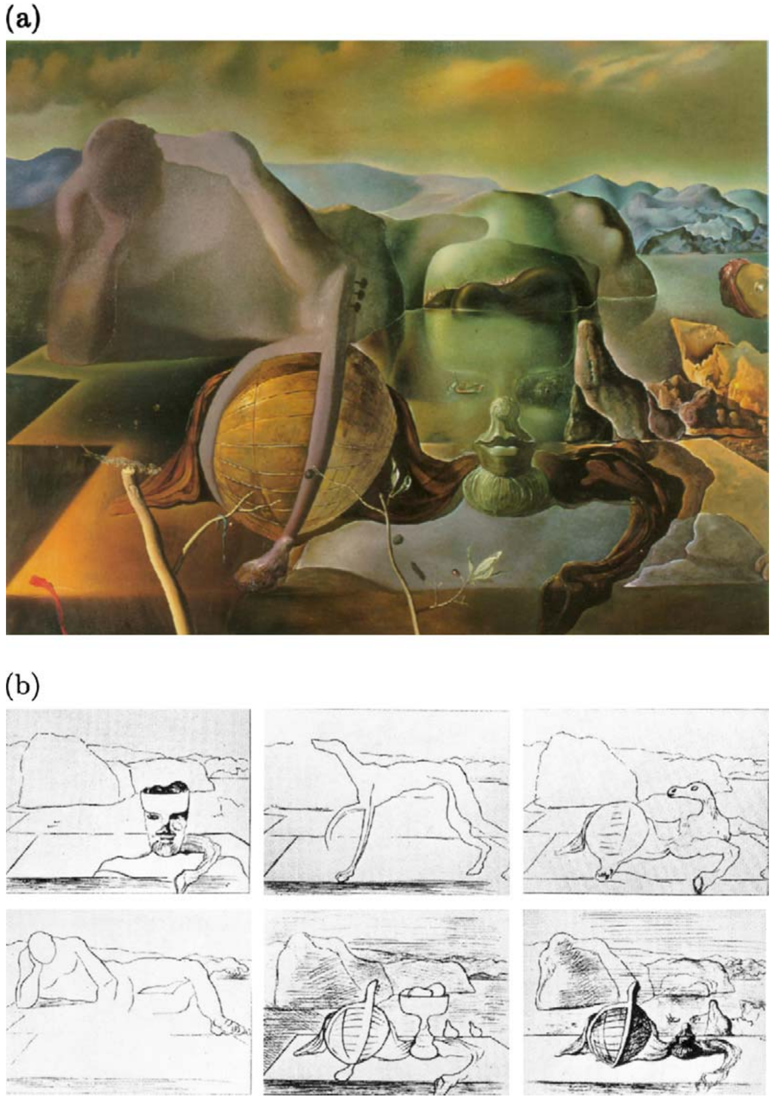


Figure 3 Dali's 'The Endless Enigma'. (a): 'The Endless Enigma', Salidor Dali, 1938, reproduced from Descharnes (1985). (b) Sketches by Dali of the images overlayed in 'The Endless Enigma'. They are (i) Face of the Cyclopean, Cretin (ii) Greyhound (iii) Mythological beast (iv) Philosopher reclining (v) Mandolin, compotier, figures on a table (vi) Woman seen from the back mending sail Descharnes, 1985)

interact with an external representation of a developing design. It is the intention of this paper to describe situated design agency in the abstract, not applied necessarily to human designers or artificial agents, so as to inform the future development of artificial agents. Descriptions inspired by human behaviour are therefore intended only to that end. We do not intend these descriptions to be taken as a cognitive model of human behaviour. Computational details in this paper apply to artificial agents but will be informal (in a computational sense); a forthcoming paper will introduce formalism to this discussion. In this paper we consider research that at first blush may seem disparate. The research reviewed in this light comes from AI, computer science, cognitive science and philosophy. The common theme is a situated, interactive approach to intelligence and problem solving. We consider these ideas in the light of the actions of human and artificial agents so as to determine what it means to say that a design agent is situated.

2 *On Agency*

We begin by clarifying what is meant by 'agent'. The definition of agency adopted here is based on that of Jennings (2000): 'An agent is an encapsulated' system 'that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives'. An agent is always situated in some environment; it receives input from its environment and acts to alter that environment. What we mean by situated will be discussed shortly. The basis of agency is autonomously acting on an environment so as to achieve the agent's objectives. We call its objectives 'goals'. Agents are differentiated from objects and other computer programs by a constraint of rationality on beliefs, goals and actions (Russell and Norvig, 2003). The agent has goals and beliefs, and executes actions based on its beliefs and goals. Deciding which actions are necessary to achieve goals is a task requiring that the agent be rational.

'Encapsulated' in the definition above means that the agent is a finite system; it is bounded in time and space. So, given this 'finitary predicament' (Cherniak, 1986), what are the conditions for rationality in an agent? The minimal conditions for an agent to be rational are given by Cherniak (1986):

If A has a particular belief-desire set, A would undertake some, but not necessarily all, of those actions that are apparently appropriate.

Similarly, the minimal inference condition concerns deductive ability:

If A has a particular belief-desire set, A would make some, but not necessarily all, of the sound inferences from the belief set that are apparently appropriate.

Without satisfying at least a minimal general rationality condition, an object cannot be an agent (Cherniak, 1986). The minimal inference condition is the basis upon which a cognitive system may be attributed to an agent. To see this, consider the alternative of ideal rationality. Ideal rationality assumes Newell's principle of rationality in an agent not be bounded in space or time:

If an agent has knowledge that one of its actions will lead to one of its goals, then the agent will select that action (Newell, 1982).

Assuming ideal rationality has important implications. It implies that an agent performs all possible inferences on all of its beliefs prior to selecting an action. A real agent is bounded and so a real agent could only accomplish this if its beliefs were limited and operated in a closed world. A real agent in an open world could only make some of the inferences possible given its current beliefs and could only undertake some of the actions resulting from such inference (Cherniak, 1986).

Autonomy is taken to mean that each agent has its own thread of control and that it decides by itself what actions to take (Jennings, 2000). A crucial difference between an agent and an object is that an object encapsulates state and behaviour realisation, but not behaviour activation or action choice (Jennings, 2000). Objects are 'totally obedient to one another and do not have autonomy over their choice of action'.

Cherniak (1986) states one further condition on agency: the minimal consistency condition:
If A has a particular belief-desire set, then if any inconsistencies arose in the belief set, A would sometimes eliminate some of them.
That is, not only should a rational agent be able to infer, it should be able to resolve inconsistencies. The condition is not that all beliefs be consistent, for such a condition would imply logical omniscience and ideal rationality. An ideal consistency condition would require that no two facts be inconsistent because classical logic says that anything can be deduced from a contradiction. The minimal consistency condition only requires that beliefs be modular and that beliefs within a module be consistent.

An ideal rationality condition also requires that an agent either

- performs all possible inferences from all possible facts, and then checks their mutual inconsistencies, or
- be designed such that it can be proved at compile time that no inconsistent facts can be asserted.

Neither of these is desirable, and in any case we may wish a design agent to temporarily assume contradictory conditions within separate modules during the early conceptual phases of designing.

3 Agency as a coupled dynamical system

Beer (1995) proposed a model of agency as a coupled dynamical system formed by an agent interacting with its environment, as shown in Figure 4.

Here, E is the environment, A is the agent, S is a sensory function and M is a motor function. S is used by Beer in a very broad sense to include any effect of the environment on the agent (and similarly for M) regardless of whether the effect is transmitted via sensors and effectors or not. Using discrete time, the agent and environment could be represented by a pair of difference equations:

$$X_{\mathsf{A}}(k+1) = \mathsf{A}(X_{\mathsf{A}}(k),\ \mathsf{S}(X_{\mathsf{E}}(k)),\ U_{\mathsf{A}})$$

$$X_{\mathsf{E}}(k+1) = \mathsf{E}(X_{\mathsf{E}}(k),\ \mathsf{M}(X_{\mathsf{A}}(k)),\ U_{\mathsf{E}})$$

Here X is a vector of state variables, U is a vector of parameters, and k is a number $1 \ldots \infty$. The first equation says that the next state of the agent is some function of the current state of the agent, the current state of the environment as sensed by the agent, and a set of parameter values. The next state of the agent depends on the environment as sensed by the agent, not on the actual state of the environment. The second equation says that the next state of the environment is some function of the current state of the environment, the current state of the agent as sensed by the environment, and a different set of parameter values. For a multi-agent system each other agent is, in this model, a part of the environment as an agent has no means of communication other than via its environment.

Each of these dynamical systems continuously deforms the flow of the other, altering each other's subsequent trajectories. The point here is that

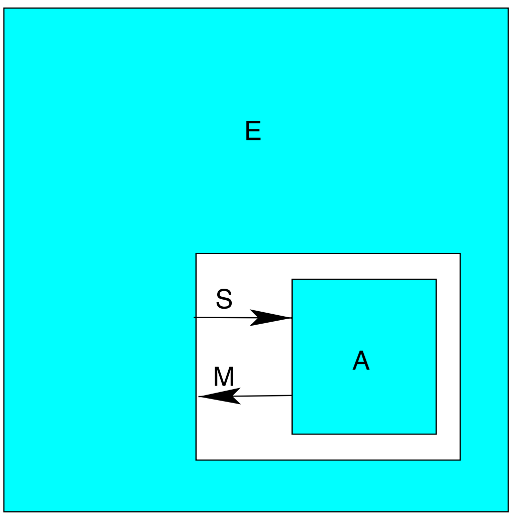


Figure 4 View of an agent and its environment as a coupled dynamical system, duplicated from Beer (1995)

'one dynamical system cannot in general completely specify the trajectory of another dynamical system to which it is coupled'
Beer, 1995

and so the agent and environment mutually perturb each other rather than having one control the other. In this view embodiment is fundamental and it is the basis of the situated view of cognition.

4 *Reasoning with indexical representations*

When I say the word 'apple', I generate in the environment a signifier that in my mind associates with a concept of an apple. This signifier is distinct from the thing that it signifies; the sound of the word 'apple' is not the same thing as either an actual apple or as the concept of an apple. This distinction between a signifier and signified is known in philosophy as the critique of meaning from structuralism (Sarup, 1993).

After the work of Peirce (Shin, 2002), signifiers are classified into three categories: indexes, icons and symbols. An index is a signifier that 'arises as a result of, or in contiguity with, the thing that it signifies' (Ashwin, 1986). Examples are shown in Figure 5. An icon is a signifier that resembles the thing that it signifies; examples are shown in Figure 6. A symbol is a signifier that does not resemble what it signifies but instead operates within agreed conventions; examples are shown in Figure 7.

(a) Footprint (b) Blackened tree (c) Tyre tracks

Figure 5 Three signifiers that are indexes. The footprint signifies use of the track by a bushwalker. The blackened tree signifies a bushfire; the lack of new green shoots signifies that it was recent. A narrow, knobbly tyre track signifies use of the track by a mountain bike; lots of narrow, knobbly tyre tracks allow us to infer that the track is popular. (a) Footprint (b) Blackened tree (c) Tyre tracks

The relationship between a signifier and the signified moves and changes with the context of its use and the experiences of the agent that uses it. This idea is important to understanding situated representations. Situated representations of the environment by a computational agent may not mean the same thing in all contexts. The idea is reminiscent of the argument put by Derrida and others that there is not a fixed one-to-one correspondence between signifiers and reality (Sarup, 1993). A post-structuralist sign is not a fixed mapping from a single referent signifying object to a single signified meaning.

An obvious example of this one-to-many correspondence is with indexicals. An indexical is an expression whose meaning is conditional on the context in which it is used. Some expressions are always objectively true: 'every frog is a frog' is true even if 'frog' is understood differently by different agents or at different times. But expressions containing indexicals are not always true and their meaning is not fixed. Examples of indexicals from the English language are first and second person pronouns, tense, time and place verbs such as 'here' and 'now' (Suchman, 1987). Consider the expression 'I saw her duck under the table' (Perry, 1997). Without knowing the context in which this expression is used we cannot say who or what 'her' signifies. Without hearing who said this we cannot say whom 'I' signifies. We also cannot say which table is being signified, and we do not know whether 'duck' signifies a verb (that someone crawls underneath a table) or a noun (an animal).

The use of indexicals is not restricted to language. Consider Figure 8. For the agent represented in the figure (by the avatar) to satisfy goals related to objects in the world requires that it associates the many

(a) Clifftop sign (b) KDE icons (c) Speed bump

Figure 6 Signifiers that are icons. The icon to the left of the clifftop sign signifies a danger of falling off the cliff. It is there because the symbolic signifier on the right of the sign (the words) require common ground on the part of the perceiver for its interpretation (that is, to understand English). The desktop icons represent the actions note, cut, lock and time. The speedbump symbol resembles the speed bump that it signifies. (a) Clifftop sign (b) Desktop icons (c) Speed bump.

(a) Rock cairn (b) Causeway signs (c) Symbolic drawing

Figure 7 Signifiers that are symbols. Each of these requires common ground on the part of depicting and perceiving agents for their understanding. The rock cairn is used by bushwalkers to signify a navigable route, which in this case is the cleft in the rock behind it. It requires common ground because inexperienced bushwalkers often do not detect these signifiers or do not understand their significance. The two signs in the centre are symbolic and presume some knowledge on the part of the perceiver, as does the drawing on the right. (a) Rock cairn (b) Causeway signs (c) Symbolic drawing

symbolic models of objects in the agent with the many objects in the environment. One way to achieve this efficiently is for reasoning by the agent to concentrate on one fixated object at a time. If the agent in the figure fixates on the lamp, that fixation point becomes its focus of attention, providing a local and exocentric three-dimensional frame of reference (Ballard et al., 1997) for perception and action. Rather than reason in terms of models of specific concrete objects, an indexical marker is set to an object that has properties of interest and further

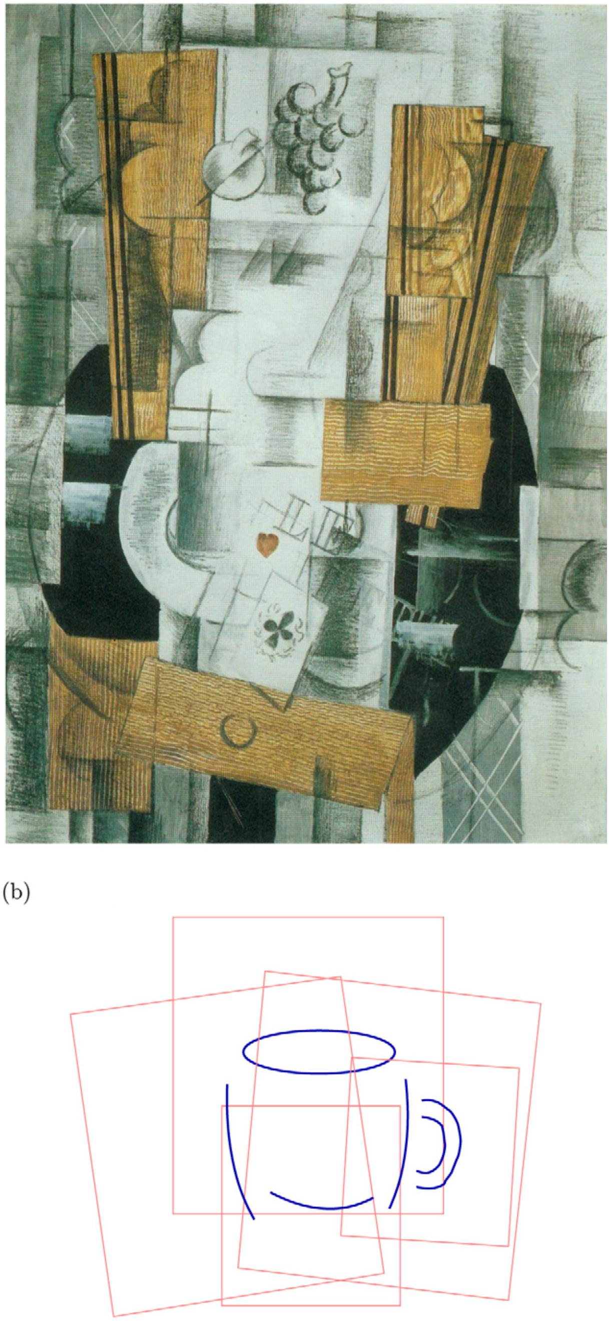
Figure 8 Agent perceiving a lamp. The agent perceives the world with respect to its own coordinate system (A_x, A_y, A_z). When it fixates on the lamp a local coordinate system (L_z, L_y, L_z) results

reasoning need only be relative to the marker. This indexical strategy is called deictic reference (Agre, 1997; Ballard et al., 1997) and it can simplify both perception and action. The robotics paradigm of interaction with an environment by actively altering the viewpoint of the robot, rather than passively observing single snapshots of it, is known as active vision. It can simplify perception by reducing the task of associating all objects with the field of view with all symbolic models to the task of associating one object at the current focus of attention with all symbolic models. It can simplify action by reducing the task of locating within the field of view all objects that have known symbolic models to the task of locating a single object.

Suppose someone looks at a cubist painting for the first time; a painting such as that shown in Figure 9(a). It is likely that the person will not perceive very much that is familiar. Suppose I then explain the Nelson and Selinger (1998) metaphor for computer vision using Figure 9(b). Nelson and Selinger approach interpreting an image not as a single, homogeneous entity but as a small number of key features with local context that are assembled in a loose global context. The metaphor is of computer vision as cubism. Viewing Figure 9(a) in this light will likely lead to different results. Interpreting a depiction is not only a matter of encoding it; we need to find what things it signifies, interpret those things, after which we may discover other signs. How we interpret the depiction varies with our knowledge, past experiences and the context in which we do the interpreting. The process of interpreting it is of focusing on one 'interesting' signifier after another and then assembling an overall interpretation.

Reasoning with indexicals is inherently situated as rules employed by the agent do not have a concrete meaning until they are employed. They are context dependent. Chapman (1989) and Ballard et al. (1997) describe tasks that illustrate the benefits of this approach. Figure 10 shows Chapman's task, which is to copy an arbitrary stack of blocks. The classical approach encodes the problem as a set of predicates like (`ON BLOCK1 BLOCK2`), (`CLEAR BLOCK3`) and searches for a set of actions that lead to another set of predicates encoding the goal. This approach is, as is shown in Chapman (1987), undecidable and intractable. (Tractable problems can be solved by algorithms where the time required on an input of size *N* is a polynomial function of *N*; problems more complex than this, such as requiring an exponentially increasing time with increasing *N*, are considered infeasible and hence intractable.) The situated alternative employs deictic markers and simple actions that operate on them. The actions do things like locate a block and set

(a)



(b)

Figure 9 Cubist images (a) Georges Braque, 'Fruit Dish and Cards', 1913, reproduced from Leymarie (1988). (b) Computer vision as cubist perception: key image fragments of a cup, duplicated from Nelson and Selinger (1998). Each fragment is perceived locally and assembled in a loose global context

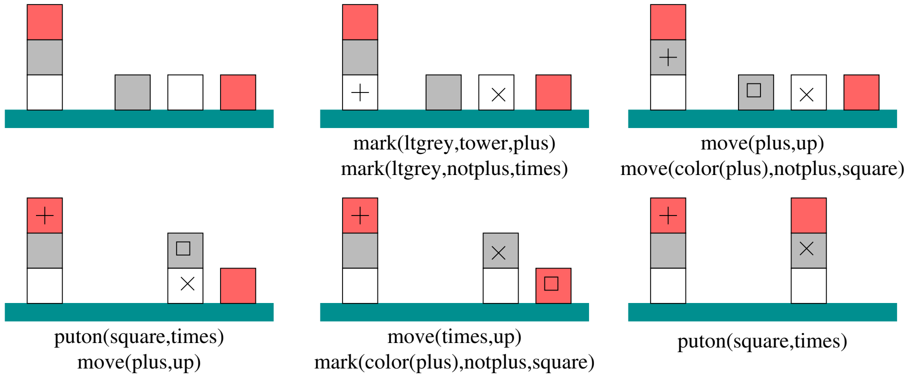


Figure 10 Chapman's (Chapman, 1989, 1991) block stacking task. The figure is derived from Ballard et al.'s (1997) drawing of it. Three deictic markers are used. The marker drawn as a plus tracks the tower being copied, the marker drawn as a times tracks the copy tower, and the third marker tracks the block that is in the copy. As can be seen, blocks are not identified uniquely; perception sets a marker to a block with requires properties (such as ltgrey *or not* ltgrey*) and after that simple marker operators complete the task*

a deictic marker on it, move the marker around, pickup a marked block and drop a marker block. Chapman (1989, 1991) shows that this simple approach can solve this kind of problem without any explicit planning, although the rules employed by the agent are dedicated to solving only one kind of problem. Ballard et al. (1997) describes similar work.

So indexicals are important to situated representation, and they provide obvious examples of cases where the signified and the signifier are not in a one-to one correspondence. Consider my pointing at an object while saying 'here it is'. If you are looking at me, can see what my finger is pointing at, do not mistake what I intend to signify because of parallax or some other error, and you know me well enough then you *may* infer a concept for what I intended to signify that is similar to what I intended. But my concept of what I intend is affected by my experiences and so it will not be identical to your concept. If you have your back turned then all bets are off. The post-structuralist contention that there is not a one-to-one correspondence between signifiers and signified objects clearly holds with indexicals. But even with non-indexicals, the interpretation of a signifier will vary with the situation in which an agent interprets it. This is because any sign is an abstraction and so can be applied to various concrete objects according to the beliefs and past experiences of the agent interpreting it and according to the context in which it is used. Icons, for example, signify things that they resemble. But 'resemble' means that at some level of abstraction they are the same, and an abstraction is on the part of the perceiving agent. Perceiving iconic signifiers requires interpretation on the part of the agent. That interpretation will vary with the experiences of the agent and will occur in the context of the current situation.

5 So, just what is 'situated'?

'Upon those who step into the same rivers different and ever different waters flow down', Heracleitus (510–480 BC), translation from Encyclopaedia Britannica (2002).

'Where you are when you do what you do matters' (Gero, 1998).

Consider the typical scenario from the ubiquitous blocks world that is shown in Figure 11. The goal of the agent shown is (ON A B) ∧ (ON B C), where (ON x y) means that the object signified by symbol x sits on top of the object signified by y, and where symbols A, B and C signify particular blocks. That is, A, B and C are not indexical as the markers in the previous section were; rather, they signify a unique object from the environment. This goal may be achieved by executing an action to move B onto C and then another action to move A onto B. An agent can plan this sequence of actions if the environment is closed and if the agent has complete knowledge of the effects of the actions. The agent can also successfully execute this planned sequence of actions if no errors occur during its activation.

Now consider Figure 12 where a second agent performs an action on the blocks unbeknownst to the first agent (this example is inspired by a similar example from Russell and Norvig (2003)). The agent can still plan its sequence of actions but as the environment has changed, the plan will no longer lead to the goal. For the first agent to succeed it either needs to control the environment to exclude all influences of which it is unaware, or it needs to continually refer back to the environment so that its reasoning does not get out of step with the environment. This is one reason why Chapman's block stacking problem is interesting: it uses simple, robust rules and deictic references so that the agent does not need to plan a complete solution. Because it continually refers to the environment before activating actions, rather than trying to plan

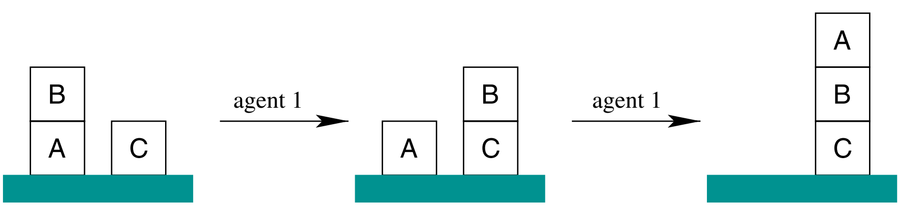


Figure 11 Blocks world scenario 1

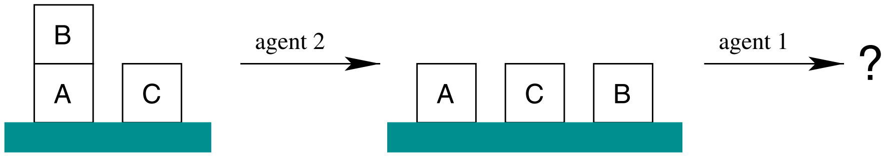


Figure 12 Blocks world scenario 2

a complete solution, it does not get out of step with that environment. The agent is situated in that environment. Rather than the agent attempting to model and then control the environment, a situated solution results that arises from the interaction between the agent and the environment.

Referring to the quote that started this section, we see that the reason we cannot step twice into the same river is that the river is constantly changing – the only unchanging thing about the river is that abstraction that is our conception of it. But an agent takes actions on a concrete environment, not an abstraction. So, if an 'agent is situated in [it's] environment' then what is *situated*? Opinions vary but there are a number of common threads. A situated approach to agency generally holds that agents are social, embodied, concrete, located, engaged and specific (Wilson and Keil, 1999). They are social in the sense of being located in a society of agents. Embodied means that actions by the agent are part of a dynamic with the world and results in immediate sensory feedback (Brooks, 1991). Concrete, located and specific mean that actions by the agent constrain its behaviour and provide a context within which it reasons and acts. Engaged means that the agent has an ongoing interaction with the environment; that planning and acting are not separated in time.

Another characteristic of a situated model relates to the character of knowledge and of its origin. Newell (1990) claims that all complex systems must be hierarchical and that with engineered systems, great care is taken to isolate a hierarchical level from those below it. Newell's knowledge level is a high level view of an agent that deliberately ignores how any knowledge is represented or acquired. It views search as the fundamental process behind intelligence, assumes an ideal unbounded rationality, and views intentions of agents as realised as plans for action that directly guide behaviour (Suchman, 1987). Situated models reject the knowledge level as an observer's model of behaviour. Clancey (1989), for instance, criticised Newell for ignoring Ryles' famous distinction between knowing that and knowing how: for conflating the capacity to perform some action with knowledge-level descriptions of it. Situated models are first person constructions that are grounded in and inseparable from environmental interaction (Clancey, 1993, 1999). In a situated view, plans are a constructed artifact of '*reasoning about* action, not the generative *mechanism of* action' (Suchman, 1987, emphasis is Suchman's). A situated view inherits the Gibsonian view that the world is its own best model (Gordon, 1997).

So if that is what we mean by *situated*, then what is a *situation*? To describe a *situation*, consider again the view described in Section 3 of an agent and its environment as a coupled system. The state of such a system at a time k is the situation $\mathfrak{S}_0(k)$. This is the state of the *system*, not just of the environment. It is also not representable by an agent any more than the Australian environment can be represented well enough for the Weather Bureau to accurately forecast the weather.

Part of the situation $\mathfrak{S}_0$ is in the agent, part is in the environment, and part arises from their interaction. That is, the whole is greater than the sum of the parts. Now, if $\mathfrak{f}_i$ is some function for entity i then $\mathfrak{S}_i(k)$ is a situation at time k for entity i:

$$\mathfrak{S}_{\mathrm{A}}(k) = \mathfrak{f}_{\mathrm{A}}(\mathfrak{S}_0(k))$$
$$\mathfrak{S}_{\mathrm{E}}(k) = \mathfrak{f}_{\mathrm{E}}(\mathfrak{S}_0(k))$$

$\mathfrak{S}_{\mathrm{A}}(k)$ is the situation at time k as represented by the agent and it depends on, but is not the same as, the current situation $\mathfrak{S}_0$. If an agent 'Sam' reasoned using a production system, say, then $\mathfrak{S}_{\mathrm{Sam}}(k)$ would depend on the set of predicates already asserted into Sam's working memory at time k, any new sense-data at time k, Sam's rules, the inference engine used, and any indirect effects of the environment on Sam (such as temperature, for example). $\mathfrak{S}_{\mathrm{A}}(k)$ is the agent's understanding of the situation at time k. This is not something that is retrieved from a database like a template (Gero et al., 2001; Smith and Gero, 2001).

'One of the most profound mistakes in the doctrine of localisation of function has been the failure to recognise that brain activity is meaningful only in a particular environmental setting.' (Rosenfield, 1988)

6 Limitations of reactive situatedness

Robots such as Brooks' do not explicitly represent and reason using goals and subgoals, but their goals are encoded implicitly in the network nevertheless. Brooks (1991) recognised that his subsumption architecture can be implemented using a production system, so his argument is over what the representations should be and not that there should be no representations whatsoever.

'I reject traditional Artificial Intelligence schemes … I reject explicit representations of goals within the machine … There can, however, be representations which are partial models of the world.' (Brooks, 1991)

Reactive systems such as are described in Brooks (1995) do make predictions about the world such as by integrating motion to predict future location, and they both plan and have goals (Brooks, 1991). But the nature of these representations is very different to symbolic (in the narrow sense) plan-execute architectures, and such agents continually refer back to the environment to update those representations.

Ginsberg (1989) defines a *universal plan* to be a function that maps situations to actions, where his definition of situation is a vector of sensor readings. He presents an argument typical of the planning community; that the expected size of a universal plan is exponential in the number of sensors. However, the term universal plan was coined by Schoppers (1989) and is objected to by Chapman (1989). The reason for the objection is that reactive systems do no explicit planning and labelling them planning gives the wrong impression. Schoppers uses the objected-to label for the following reason. Consider a production system implementation of a plan. This may employ a large number of features in a large number of rules, but as there are many practical production system implementations it is clearly feasible. First, variables in the rules act indexically so that each rule applies in a range of situations. In this way they abstract large parts of the underlying space to the indexical variables and allow other parts to be ignored. Secondly, these reactive rules can act recursively so that three-block tower building rules can build towers of any height, and the indexical variables allow the types of towers that are built to change.

Solving a jigsaw puzzle is an example of a task that is suited to situation-determined behaviour (Kirsh, 1996) as they are perceptually hard but conceptually simple. This is because only a few tiles need be considered at any time and the possible actions are constrained by the perception of matching tile patterns. Interpreting Figure 3 is also solving a puzzle (Elkins, 1999), and we know that Dali intended it to be interpreted so from his sketches shown as Figure 13. Designing is not necessarily like solving a puzzle but the interaction inherent in solving these kinds of puzzles have clearly analogs to that in some kinds of designing. Minsky's (1986) 9 dots puzzle of Figure 3(b) is an example where a problem is discovered perceptually but realising the solution requires something more. What is also required are the abilities to frame and test hypotheses, and to learn. An agent that designs by interacting with a sketch needs good situated perception and action processes, but it also must be able to link its interpretations of those sketches to concepts that signify design requirements. While Brooks validly criticises Kirsh (1996) and others for misreading his theory as requiring intelligence 'without representation' whatever, his critics do have a valid point about the probable limits of purely reactive agents. This is especially apparent with design agents, where the goals of the agent are to satisfy what are initially likely to be quite abstract, conceptual requirements.

7 *On constructive memory*

'... whilst part of what we perceive comes through our senses from the objects before us, another part (and it may be the larger part) always comes out of our head.' (Gordon, 1997, quoting William James, 1890)

'All action is embodied because perception and action arise together automatically: Learning is inherently "situated" because every new activation is part of an ongoing perception–action coordination' (Clancey, 1993).

If the beliefs of an agent are situated, then the application to design problems should also be situated. The memory of an agent reflects how it has adapted to its environment. So if an agent is to design autonomously then it should apply its learned beliefs as a design evolves. Learning is 'the improvement of performance in some environment through the acquisition of knowledge resulting from some experience in that environment' (Langley, 1996). As it involves an improvement of performance, learning by an agent occurs with respect to some task and environment.

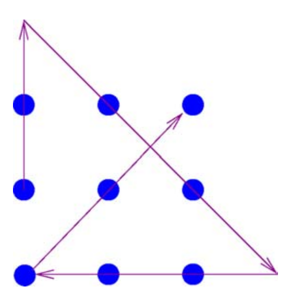


Figure 13 The 9 dots puzzle from Minsky (1986). The task is to join the dots with only four straight lines without lifting the pen. Most people given this for the first time assume a rule that does not exist: that the lines cannot go beyond the bounds of the dots

At the lowest level, an agent receives sensor signals. These are uninterpreted and, as the world does not come pre-organised into categories, a process of assimilation constructs percepts out of these sensor signals according to the agent's current goals and expectations. 'A cognitive organism perceives (assimilates) only what it can fit into the structures it already has' (von Glasersfeld, 1995). From an external observer's point of view, perception modifies what is perceived in order to fit into the agent's existing conceptual structures, but the agent itself recognises no modification. If all representations used by an agent are provided by an external observer, 'learning' becomes a form of search technique. For genuine learning to occur, an autonomous agent requires the ability to construct fundamentally new models (Morrison, 1997). The agent can learn by interacting with the environment, performing actions and receiving percepts via its own sensors, comparing predictions of existing models against received percepts, and adapting existing models accordingly. The result is representations by the agent that are grounded in its experiences.

Realists believe that reality exists and does so independently of anyone's experience of it (Hoffman, 1994). An important question, though, is how it is that we come to know reality and how is it that we represent it. Constructivists may accept that there is a 'reality' but do not accept that we are directly aware of it. The idea is of the perceiver of an attended object as being like the captain of a submarine (Gordon, 1997): they have knowledge of the medium in which they are submerged but they cannot experience it directly.

A constructivist ontology holds that there is no isomorphism between an object in the 'real' environment and an internal model of an agent that perceives it (von Glasersfeld, 1995). The agent does not *directly* perceive the environment and so cannot learn cause and effect from observation (Armstrong, 1988). There can be no cause and effect between environmental objects and internal conceptual objects.

'Any notion that cognitive structures could come to reflect ontological reality – e.g., that we could discover the ontic shape of things by sliding our senses or measuring instruments along the surfaces of things-in-themselves and thus plot deliberate contacts – is an illusion.' (von Glasersfeld, 1995)

Because of this there is inference on our part between the sense-data from an attended object and our conscious experience. Constructive models of perception view this inference as being hypothesis driven. Perceptions can be noisy or ambiguous (such as Figure 14) and even paradoxical (such as Figure 15). The way that this is handled is by processing percepts both top down and bottom up.

The distinction in epistemology between knowledge and belief is that knowledge is justified true belief. Now in a constructive model of agency any beliefs are constructed by the agent and cannot be directly verified by comparison with the environment (Hoffman, 1994). If this is the case then constructivist agents only have beliefs. Those beliefs cannot be compared against an objective ‘reality’ and so an agent never knowingly has any knowledge.

If only a part of a person’s conscious experience of an object is actually what was sensed, their memory of that experience at a later time will largely be a memory of an inference. If perception is constructed then a memory of it will also be an inference, at least in part. This is the basis of the theory of constructive memory. Recall of an experience is a synthesising process in which ‘what must have occurred’ is reconstructed from a few fragments in working memory (Cherniak, 1986). It is in contrast to a filing cabinet view of memory; the traditional AI view in which symbols are received from the environment and stored for later use.

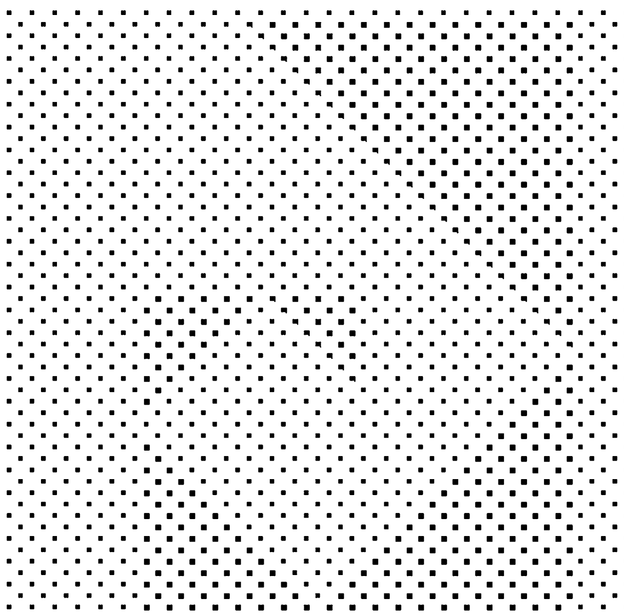


Figure 14 KCDCC emergent dot logo. This can be viewed online, to better effect, at http://www.arch.usyd.edu.au/~g_smith/internetProg/jsDots/jsdots.html. The effect changes with viewing distance

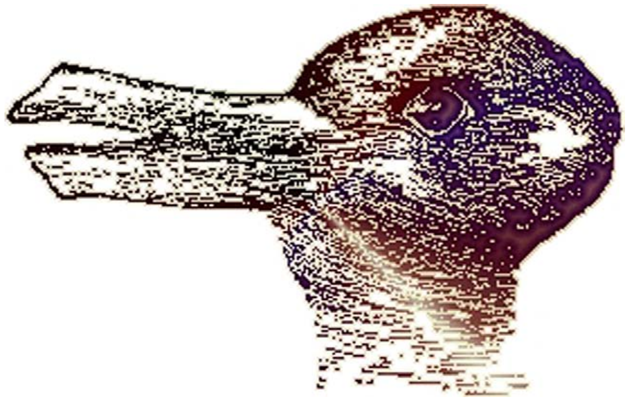


Figure 15 Joseph Jastrow 1900 Duck-Rabbit Drawing, from http://members.lycos.nl/amazingart/E/10.html

Classically, a logician selects a syntax, semantics and rules of inference and these define a landscape of the truth of all expressible sentences. Performing inference is a matter of searching for a path through that landscape from the starting location to the goal location: the truth of the goal location already exists, the agent just has to find it. So some term p already exists in the truth landscape and even if its value is not known, because it already exists it can be guaranteed that either p is true or $\neg p$ is true. In a constructivist view the logical landscape is not automatically defined when the logic is defined. It is more like creating a new virtual world; the kinds of world that you can create are constrained but initially the world is empty. It is not that a logician (or agent) searches for a path through an existing landscape; they actually construct the truth landscape as they go.

A constructive memory reflects how cognitive systems have adapted to their environment. Recollection involves a reconstruction of past experiences based on information in the current environment and on the way cognitive processing is currently accomplished – more like fantasising than looking up records (Guenther, 1998). The theory predicts that people will not be able to remember very well the constantly changing details of events, but it should be easy for people to remember the invariants or enduring patterns of events (Guenther, 1998). AI theories that are potentially constructivist include neural networks (provided they are unsupervised and learn interactively), Drescher's (1991) Piagetean schemas and the genetic classifier systems of Holland (1986).

Constructivist learning is through assimilation and accommodation (von Glasersfeld, 1995). Because perception depends on inference by the agent, interaction by the agent with its environment depends on its expectations. Sense-data, therefore, are attempted to be assimilated into existing beliefs and expectations. Learning results when the agent is unable to do this, with new characteristics revealed that were previously disregarded. This learning from unexpected results is accommodation.

Theories of constructive perception and constructive memory deal with how an individual agent interacts with its environment. In such cases there are no objectively valid beliefs. But considering either human societies or multi-agent systems requires that beliefs about communication be accounted for. How is that two agents viewing the same signifier come to believe that they agree on the meaning of that signifier? Recent work on social constructivism by Ernest (1998) and others has extended earlier work by explicitly recognising the role of communication between agents. Social constructivism holds that agreements on what signifiers and communicated symbols mean develop autonomously with social interaction. The meaning of a mathematical symbol or axiom is not defined a priori; it means whatever the community of mathematicians come to agree it means (Ernest, 1998). The same may be said of designers and technical drawings.

8 Schön and the dialectic of designing (Schön, 1987; Schön and Wiggins, 1992)

The modern meaning of dialectic[1] is to maintain something through a process of thesis, antithesis and dialectic synthesis (Popper, 1996b). This three-part process is used to describe both scientific and philosophical inquiry, but it can also be used to describe the process of designing. Thesis is a target idea; with designing it is behaviour interpreted from the current design. Antitheses are flaws or errors found with the idea; with designing they are expected behaviours that remain unsatisfied. There is an ongoing tension between thesis and antithesis that the dialectic synthesis takes steps to resolve.

A dialectic is not simply trial-and-error search (Popper, 1996b). While trial-and-error eliminates a candidate solution at each step, dialectic is focused on building upon what already exists. Some part of what already exists is preserved at each step, and the synthesis updates some part of what was found by antithesis to be at fault. The tension that is resolved by the dialectic synthesis are contradictions between thesis and antithesis. But these are not logical contradictions in the sense of classical logics; rather, they are properties required by antithesis that cannot be constructed by the thesis. The role of a dialectic synthesis is to ensure that these missing properties can be constructed.

Now, designers bring to their practice a repertoire of routinised responses that are spontaneous, tacit and not conscious (Schön, 1985). Indeed, this is characteristic of human behaviour. Schön calls this *knowledge-in-action*: attentive action that is largely spontaneous and automatic. Sketching and talking are examples. Applying this tacit knowledge results in consequences that are as expected. Occasionally, though, some consequences will not be as expected. If an unexpected consequence is noticed, reflection on the surprising consequence results by the designer. They consciously reason about the surprise, but they must do so while continuing to design if such reasoning is to have an effect on the target design. Schön calls reflecting on the outcomes of an activity whilst continuing to perform that activity *reflection-in-action* (Schön, 1985, 1987; Schön and Wiggins, 1992).

A dialectic view of designing is of experimenting, but unlike scientific experimenting it deals with worlds that do not yet exist. Design experiments occur 'on the spot' (Schön, 1985) as experiment and analysis are not separated in time as with scientific experiments are. Designing is a 'conversation with materials conducted in the medium of drawing' (Schön and Wiggins, 1992) and as such depends on both perception and action.

Popper (1996a) argued that 'science should be visualized as progressing from problems to problems', and designing progresses in the same way (Bamford, 2002). The unit of design is not a design concept but instead is an action (Dorst and Dijkhuis, 1995). A designer actively structures a problem, evaluating his/her own actions whilst restructuring and solving a problem. In considering this, Dorst and Dijkhuis (1995) find that a problem solving paradigm is apt where the problem is 'fairly clear cut', but that reflection-in-action is better suited to conceptual design. The process is one of a designer interacting with an external representation, performing a series of experiments the results of which influence further experiments. It is a reflective conversation with the situation (Clancey, 1997). Each design action is perceived globally and across all design domains however. The reason for this approach is that attempting to consider the implications of a move across all domains simultaneously is equivalent to searching many non-isomorphic graphs simultaneously.

Instead, in selecting the next action by considering one design domain the task of selecting actions is made tractable and the process is transformed into the design experimentation described. The designer synthesises a potential design action, and performs a design experiment by sketching and interpretation. Expectations by the designer of the impact of the action are the dialectic thesis and are formed with respect to the design domain that was used to synthesise the action. The dialectic antithesis is the evaluation of the current design across all design domains and with respect to design requirements that have not yet been satisfied. Computationally this is like rather more like Agre's (1997) dynamic difference reduction than search: find something to do, do it, and find something else to do. The effect is to treat the external representation of a developing design as a kind of multiple representation. It is one in which, unlike internal symbolic multiple representations, all of the representations of the domains are automatically synchronised at each design action.

9 *Conclusion*

The problem and the solution to many designing tasks emerge together from an interaction between a situated designer and the medium of the design. So the question of whether a situated model of design is useful and appropriate to design is now being asked. Is it useful and appropriate for modelling artificial agents or only for describing human designers? Is conceptual design an activity which is inherently best viewed as situated interaction with an external representation, or is it only a psychologically appropriate model? These are questions towards which research effort is now being expended.

Answering these questions requires there being a firm understanding of what these terms mean, especially the term 'situated', in a design context. Describing agents as 'situated' does not clarify our understanding of designing described in this way if we approach it with different understandings of what it means for a design agent to be 'situated'.

This paper provides an answer to the question of what it means when we say that a design agent is situated. We reviewed notions of agent-ness, rationality, and indexicality. Applying these notions allow us to say what a situation and what situatedness are. Applying these notions to embodied action and constructive memory allow us to consider how such representations may be acquired and constructed.

Acknowledgements

This work was supported in part by a University of Sydney Sesqui R&D grant and an Australian Postgraduate Award at the University of Sydney, Australia.

References

Agre, P E (1997) *Computation and human experience* Cambridge University Press

Armstrong, D M (1988) *Perception and belief Perceptual knowledge, Oxford readings in philosophy* pp 127–144

Ashwin, C (1986) Drawing, design and semiotics in **V Margolin** (ed) *Design discourse: history, theory, criticism*, University of Chicago Press pp 199–209

Ballard, D H, Mayhoe, M M, Pook, P K and Rao, R P N (1997) Diectic codes for the embodiment of cognition *Behavioral and Brain Sciences* Vol 20 No 4 pp 723–742

Bamford, G (2002) From analysis/synthesis to conjecture/analysis: a review of Karl Popper's influence on design methodology in architecture *Design Studies* Vol 23 No 3 pp 245–261

Beer, R D (1995) A dynamical systems perspective on agent–environment interaction *Artificial Intelligence* Vol 72 pp 173–215

Brooks, R A (1991) Intelligence without reason in *Proceedings of the 12th International Conference on Artificial Intelligence* pp 569–595

Brooks, R A (1995) Intelligence without representation in **G F Luger** (ed) *Computation and intelligence*, MIT Press pp 343–362

Chapman, D (1987) Planning for conjunctive goals *Artificial Intelligence* Vol 32 No 3 pp 333–377

Chapman, D (1989) Penguins can make cake *AI Magazine* Vol 10 No 4 pp 45–50

Chapman, D (1991) *Vision, instruction and action* MIT Press

Cherniak, C (1986) *Minimal rationality* MIT Press

Clancey, W J (1989) The knowledge level reinterpreted: modeling how systems interact *Machine Learning* Vol 4 pp 285–291

Clancey, W J (1993) Situated action: a neurobiological interpretation response to Vera and Simon *Cognitive Science* Vol 17 No 1 pp 87–116

Clancey, W J (1997) *Situated cognition: on human knowledge and computer representations* Cambridge University Press

Clancey, W J (1999) *Conceptual coordination: how the mind orders experience in time* Lawrence Erlbaum

Coyne, R D, Rosenman, M A, Radford, A D, Balachandran, M and Gero, J S (1990) *Knowledge-based design systems* Addison-Wesley

Descharnes, R (1985) *Dali* Thames and Hudson

Dorst, K and Dijkhuis, J (1995) Comparing paradigms for describing design activity *Design Studies* Vol 16 pp 261–274

Drescher, G L (1991) *Made up minds* MIT Press

Elkins, J (1999) *Why are our pictures puzzles? On the modern origins of pictorial complexity* Routledge

Encyclopaedia Britannica (2002) *Heracleitus The New Encyclopaedia Britannica* (15th edn)

Ernest, P (1998) *Social constructivism as a philosophy of mathematics* State University of New York Press

Gero, J S (1998) Conceptual designing as a sequence of situated acts in **I Smith** (ed) *Artificial intelligence in structural engineering*, Springer pp 165–177

Gero, J, Saunders, R, Smith, G, Liew, P, Jupp, J (2001) How situations are constructed, notes for meeting on 23 may 2001 (4:30 pm), Key Centre of Design Computing and Cognition, University of Sydney. Available at: http://www.arch.usyd.edu.au/kcdc/situatedness/

Ginsberg, M L (1989) Universal planning: an (almost) universally bad idea *AI Magazine* Vol 10 No 4 pp 40–44

Gordon, I E (1997) *Theories of visual perception* John Wiley and Sons

Guenther, R K (1998) *Human cognition* Prentice-Hall

Hoffman, R R (1994) Constructivism versus realism, or constructivism and realism *Journal of Experimental and Theoretical Artificial Intelligence* Vol 6 pp 431–435

Holland, J H (1986) Escaping brittleness: the possibilities of general purpose learning algorithms applied to parallel rule-based systems in **R S Michalski**, **J S Carbonell and T M Mitchell** (eds) *Machine Learning II*, Kaufmann pp 593–623

Horgan, J (2004) The end of science revisited *Computer* Vol 37 No 1 pp 37–43

Jennings, N R (2000) On agent-based software engineering *Artificial Intelligence* Vol 117 pp 277–296

Kirsh, D (1996) Today the earwig, tomorrow man? in **M A Boden** (ed) *The philosophy of artificial life, Oxford readings in philosophy*, Oxford University Press pp 237–261

Langley, P (1996) *Elements of machine learning* Morgan Kaufmann

Lawson, B (1997) *How designers think: the design process demystified* (3rd edn) Architectural Press

Leymarie, J (1988) *Georges Braque* Prestel-Verlag

Margolin, V (1986) *Design discourse: history, theory, criticism* University of Chicago Press

Milligan, S and Shand, N (1996) *The best of Q*

Minsky, M (1986) *The society of mind* Simon and Schuster

Morrison, C T (1997) Are we wrong about representation? *Journal of Experimental and Theoretical Artificial Intelligence* Vol 9 No 4 pp 441–469

Nelson, R C and Selinger, A (1998) A cubist approach to object recognition in *International Conference on Computer Vision (ICCV98)*. Bombay, India pp 614–621 Available atftp://ftp.cs.rochester.edu/pub/u/nelson/1998_iccv_ext.ps.gz

Newell, A (1982) The knowledge level *Artificial Intelligence* Vol 19 No 1 pp 87–127

Newell, A (1990) *Unified theories of cognition* Harvard University Press

Perry, J (1997) Indexicals and demonstratives in **B Hale and C Wright** (eds) *A companion to the philosophy of language*, Blackwell pp 586–612

Popper, K R (1996a) *Truth, rationality, and the growth of scientific knowledge Conjectures and refutations: the growth of scientific knowledge* (5th edn) Routledge pp 215–250

Popper, K R (1996b) *What is dialectic? Conjectures and refutations: the growth of scientific knowledge* (5th edn) Routledge pp 312–335

Rosenfield, I (1988) *The invention of memory: a new view of the brain* Basic Books

Russell, S and Norvig, P (2003) *Artificial intelligence: a modern approach* (2nd edn) Prentice Hall

Sarup, M (1993) *An introductory guide to post-structuralism and post-modernism* University of Georgia Press

Schön, D A (1985) *The design studion: an exploration of its traditions and potentials* RIBA Publications

Schön, D A (1987) *Educating the reflective practitioner* Jossey-Bass

Schön, D A and Wiggins, D (1992) Kinds of seeing and their functions in designing *Design Studies* Vol 13 No 2 pp 135–156

Schoppers, M J (1989) In defense of reaction plans as caches *AI Magazine* Vol 10 No 4 pp 51–60

Shin, S-J (2002) *The iconic logic of Peirce's graphs* MIT Press

Smith, G J and Gero, J S (2001) Emerging strategic knowledge while learning to interpret shapes in **J S Gero and K Hori** (eds) *Strategic knowledge and concept formation III, key centre of design computing and cognition*, University of Sydney pp 69–86

Suchman, L A (1987) *Plans and situated actions* Cambridge University Press

von Glasersfeld, E (1995) *Radical constructivism: a way of knowing and learning* The Falmer Press

Wilson, R A and Keil, F C (1999) *The MIT encyclopedia of the cognitive sciences* MIT Press

1. The ancient Greek meaning was restricted to an argumentative use of language (Popper, 1996b).