

From extended mind to collective mind

Action editors: Luca Tummolini and Cristiano Castelfranchi

Deborah Perron Tollefsen *

Department of Philosophy, University of Memphis, 327 Clement Hall, Memphis, TN 38152, USA

Received 30 March 2005; accepted 7 November 2005

Available online 28 February 2006

Abstract

Although the notion of collective intentionality has received considerable attention over the past decade, accounts of collective belief and intention remain individualistic. Most accounts analyze group intentional states in terms of a complex set of individual intentional states and, thus, it is individuals not groups that *have* intentional states. In this paper, I attempt to undermine one of the motivations for refusing to acknowledge groups as the bearers of mental states. The resistance to collective mental states is motivated by the view that mental states are located in minds and minds are in heads. Since groups do not have heads or brains, they cannot have minds or mental states. There is a significant and important thesis in cognitive science, however, which suggests that the mind is not bounded by skin and bones. If “the mind ain’t in the head”, then this removes a major barrier to the idea of collective minds.

© 2006 Elsevier B.V. All rights reserved.

Keywords: Extended mind; Collective mind; Collective intentionality

1. Introduction

Until recently, “collective intentionality” was a phrase that raised skeptical eyebrows in many circles. A growing number of philosophers and researchers in the social and cognitive sciences, however, have begun to take the idea of collective intentionality very seriously. Despite the growing interest, standard accounts of collective intentional states preserve a form of individualism. Collective beliefs and intentions on these accounts are not the mental states of some group agent but are to be identified with a complex set of individual intentional states.¹

In this paper, I attempt to undermine one of the motivations for refusing to acknowledge groups as the bearers of mental states. The resistance to collective mental states is motivated by the view that mental states are located in minds and minds are located in heads. Since groups do not have heads or brains, they cannot have mental states. There is a significant and important thesis in cognitive science, however, which suggests that the mind is not bounded by skin and bones. If “the mind ain’t in the head”,² then this removes a major barrier preventing acceptance of the idea that groups are the bearers of mental states. Having removed one of the major theoretical objections to the idea of collective minds, I take up the pragmatic issue of the explanatory power of the collective mind hypothesis.

In Section 2, I focus on the work of Andy Clark and David Chalmers. Although there have been a variety of recent attempts to stretch the boundaries of the mind, Clark and Chalmers (henceforth, C and C) provide the

* Tel.: +1 901 678 4689; fax: +1 901 678 4365.

E-mail address: dtollfsn@memphis.edu.

¹ For instance, see Bratman (1993) and Tuomela (1992). Searle (1990, 1995) also preserves a form of individualism by claiming that we-intentions and we-beliefs are individual mental states. It is less clear that Gilbert (2002, 2003) is committed to this form of individualism. Plural subjects are the appropriate target of attitude ascription. At times, however, she suggests that the beliefs of a group are only analogous to the beliefs of individuals. This suggests that she does not view groups as *literally* having beliefs.

² The phrase appears in McDowell (1992) and is a variation on Putnam’s “meaning ain’t in the head” found in (1975).

clearest and most sustained argument for the extended mind hypothesis. In *The Extended Mind* (Chalmers & Clark, 1998) and in many recent articles and books by Clark, they advocate what they call *active externalism*.³ Active externalism is the view that aspects of an individual's environment, to which the individual is linked in a two-way interaction, are as much a part of human cognition as are other parts of the human brain. Computers, calculators, palm pilots, even post-it notes, are artifacts which individuals use in their cognitive endeavors. The interaction between these artifacts and the individual constitutes a *coupled system* that functions as a cognitive system in its own right. Thus, C and C argue that cognition extends beyond the skin and with it the mental states that support cognition. In some cases, *beliefs* are constituted partly by features of the environment, when those features play the right sort of role in the cognitive process. The focus of C and C's work has been on coupled systems that involve a single individual and an artifact, such as a computer. I will call such couplings – *solipsistic systems* – to mark that they do not involve other agents. In Section 3, I extend the C and C argument to establish the possibility and plausibility of *collective systems*, coupled systems that are constituted primarily by humans. In Section 4, I consider many of the objections against active externalism and show that they lose their force when we consider the possibility of collective systems. In Section 5, I conclude by responding to one of the few sustained attacks on the explanatory power of the collective mind hypothesis.

2. The extended mind

C and C argue for active externalism on the basis of several thought experiments. The first involves a video game that was popular during the 1980s. Tetris involves manipulating falling objects in order to make them fit into an arrangement of objects on the bottom of the screen. As the objects begin falling at an increased speed, the task becomes more difficult. There are two options for the manipulation of the objects: (1) mentally rotate the shapes in order to figure out where they might be placed or (2) use the control button that causes the falling objects to rotate in various ways and make the assessment of fit based on what is seen. C and C ask us to entertain the following additional option. Imagine that (3) sometime in the future a rotation device similar to the one on the computer game is implanted in the brain. The device rotates an image of the object on demand and would be just as quick and easy as using the rotation button. We would simply issue some sort of mental command and the rotation would be completed just as if we had manipulated the control button with our hands.

C and C now ask us to consider the differences between these cases. Case 1 is clearly an instance of mental rotation. It appears to be a paradigm case of mental cognition. Intuitions suggest that case 2 is non-mental. After all, the rotation does not occur “inside the mind”. What about the case involving the prosthetic rotation device? By stipulation the device does exactly what the device in 2 does. The only difference is that it is lodged in the head. Is this really a difference that matters? If we found an alien species that developed a device like that in 3 via natural selection we would, it seems, have no difficulty counting it as a case of mental rotation. So, what then, prevents us from saying that case 2 is a case of mental cognition? Is the mere fact that the device is outside the head enough to exclude it from the realm of the mental? It seems not. C and C offer the following principle:

Parity principle: if, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting it as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process. (1998, p. 644)

The use of notebooks, palm pilots, calculators, and other artifacts are not simply tools for aiding our cognitive endeavors. In some cases, they are functionally equivalent to mechanisms like short-term and long-term memory, mental images, mental calculations, and so on. We would have no problem accepting them as part of the cognitive process if they were located in the head and so, according to the parity principle, these devices ought to be considered part of the cognitive process of a system that includes both human body and environment.

C and C are quick to note that there are, of course, differences. One obvious difference is that cognitive processes are transportable and easily accessible. Our ability to mentally rotate an image and our imagined alien's ability to rotate can be easily accessed and toted around to manipulate various other geometrical shapes and solve various other problems. This is not the case with the rotation device attached to the Tetris console. This suggests that it is not geography that distinguishes internal processes as mental, but the properties of portability and accessibility.

C and C agree and in order to respond to this worry and to show that it is not just cognition but states like belief that are extended, they offer this second thought experiment. Consider Inga and Otto. Inga hears about the minimalist exhibit at the museum of Modern art in New York (MOMA). She recalls that it is on 53rd street and based on her memory she begins her journey to the museum. Otto has a mild form of Alzheimer's. He uses a notebook to record phone numbers, addresses, dates, names, and so on. Because his memory is so poor he carries this notebook with him at all times. When he hears about the minimalist exhibit he pulls out his notebook

³ Clarke has defended and developed the view in several books (2001a, 2003) and articles (2001b, 2004a).

and looks up the address and then sets out to visit MOMA. Now what is the difference between Otto's use of the notebook and Inga's use of her memory? C and C claim nothing. We would have no trouble explaining Inga's behavior by appeal to her desire to attend the exhibit and the belief that the museum is on 53rd street. Likewise, we should have no trouble explaining Otto's behavior by appealing to the same desire and belief. The only difference is that in Otto's case the belief is stored in his notebook rather than his biological memory (1998).

Certainly, insofar as beliefs and desires are characterized by their explanatory roles, Otto's and Inga's cases seem to be on par: the essential causal dynamics of the two cases mirror each other precisely... The moral is that when it comes to belief, there is nothing sacred about skull and skin. What makes some information count as a belief is the role it plays, and there is no reason why the relevant role can be played only from inside the body.

It is important to note here that C and C allow for the possibility that the content of occurrent mental states supervene locally, that is, on processes inside the brain. The Otto and Inga cases, however, are cases of dispositional long-term beliefs. C and C argue that the physical vehicles of such non-conscious mental states have a supervenience base that extends beyond the skin and into the environment.⁴

One might try to distinguish Otto and Inga by saying that all Otto actually believes is that the notebook has the address and that this in turn leads him to look into the notebook and form the *new* belief about the actual street address. This story is what Clark calls the Otto-2 step (Clark, 2004b, p. 7). Clark quickly dismisses this. Note that we can run the same 2-step on Inga. All Inga really believes is that the information is stored in her memory. When she retrieves the information she forms the *new* belief about the street address of the MOMA. But we do not run this story. Indeed, it is a bizarre account of how beliefs in long-term memory explain our behavior. It is highly unlikely that Inga has any beliefs about her long-term memory. As Clark puts it,

She just uses it, transparently as it were. But ditto for Otto: Otto is so used to using the book that he accesses it automatically when bio-memory fails. It is transparent equipment for him just as the biological memory is for Inga. And in each case, it adds needless and psychologically unreal complexity to introduce additional beliefs about the book or biological memory into the explanatory equations (Clark, 2004a, p. 8).

But does this mean that every time we use a computer, rely on a pen and paper to compute a sum, or look up a phone number we form a coupled system? Does my mind extend to encompass the phone book on my phone table? In order to address these worries C and C offer the following criteria to be met by non-biological candidates for inclusion into a coupled system:

1. 'The resource(s) must be available and typically invoked' (Clark, 2004a, p. 6). Otto always uses his notebook and carries it with him. He appeals to it on a regular basis when asked questions such as "Do you know...?"
2. 'That any information thus retrieved be more or less automatically endorsed. It is not always subject to scrutiny. It should be deemed as trustworthy as something retrieved clearly from biological memory' (Clark, 2004a, p. 6).
3. 'Information contained in the resource should be easily accessible' (Clark, 2004a, p. 7).
4. Finally, to avoid some obvious objections involving readily available books and internet search engines, the information contained in the resource must have been previously endorsed by the subject. It is Otto who places the information in his notebook. If it just appeared there we would probably not grant it the same status as that of a belief (1998).

C and C's argument is controversial. The hypothesis that the mind and environment sometimes form a coupled system which is the locus of cognition and belief raises some puzzling issues concerning the self, identity, agency and responsibility. For now, I want to grant C and C their conclusion. The mind extends beyond the skin to encompass resources in the environment. In the next section I argue that these resources are not only non-biological artifacts but other biological agents. When minds extend to encompass other minds, there is a collective system formed. This is a possibility that C and C gesture at but do not fully develop.⁵

My approach may seem philosophically imprudent. To support one's theory by extending an already controversial argument seems risky. But, as we shall see, C and C's argument is not easily dismissed and because the idea of

⁴ The distinction between vehicle and content will help to distinguish between active externalism and the content externalism advocated by Putnam (1975), Burge (1979), and others. Content externalism holds that the content of a mental state is determined by environmental or causal factors. My twin on twin earth does not have water beliefs because Twin earth contains XYZ not H₂O. The content differs. The vehicle, however, remains, in the original thought experiments involving duplicates, the same and inside the head. Active externalism argues that the vehicle of content need not be restricted to the inner biological realm. The idea is that both cognitive contents and cognitive operations can be instantiated and supported by both biological and non-biological structures and processes.

⁵ It is also present in the work of social psychologist/cognitive scientist Edwin Hutchins (1995) but he does not develop the philosophical issues and address the philosophical problems related to this idea.

collective systems resists many of the objections formulated against the extended mind, it turns out that extending this controversial argument in the way that I suggest makes my conclusion less controversial.

3. From solipsistic systems to collective systems

Let us begin with the Tetris thought experiment. Recall, the three options:

- (1) mental rotation;
- (2) rotation by using the control button on the console;
- (3) rotation by device implanted in the brain.

C and C argue that because we would count (1) and (3) as a case of cognition and the only difference between (2) and (3) is geography, (2) is a form of “mental” cognition. Consider a fourth option: (4) rotation by issuing a command to one’s friend who rotates the image for you using the control button on the console. (4) is not a very efficient way of playing Tetris. Options (1)–(3) are clearly faster forms of cognition. But if (2) counts as cognition, then how can we rule out (4)? If C and C are correct, the mind can extend to encompass other individuals. My Tetris friend and I form a coupled system and cognition (of a limited sort) is distributed over this system.

One might try to rule (4) out by pointing to the fact that when another individual enters the scenario there is a significant difference in control. I can control the button on the computer in a way that I cannot control my friend. My friend can simply refuse to rotate. Computers, assuming they are functioning well, cannot just decide to opt out of the coupled system. But surely our reliance on the functioning of the computer is just as precarious as our reliance on the functioning of a friend. The reason for their malfunction may differ (the computer battery may die, the friend may get annoyed) but in both cases our control of the rotation is somewhat dependent on something outside of us.

But still, one might argue that my Tetris friend lacks the same sort of portability and accessibility that the Tetris console lacks. I might occasionally play Tetris with my friend but he is not a ‘constant’ in my life, as required by the criteria C and C established. So let us consider again the case of Otto.

Otto, according to C and C, forms a coupled system with his notebook. Otto’s notebook functions in the same way that his long-term memory functions and according to the parity principle should be considered part of Otto’s mind. Now consider Olaf who is married to Inga. They have been married for 30 years. Olaf does not suffer from Alzheimer disease. He is, however, a philosopher. He often gets lost in his work and has difficulty remembering his appointments, phone numbers, addresses, and so on. Inga, however, has a sharp mind and because they spend a great deal of time together Inga provides Olaf with all

of the information that he needs in order to get through his day. Indeed, Inga seems to serve just the same purpose that Otto’s notebook serves him. She is his external memory. Does this mean that Olaf’s mind extends into Inga? Do Otto and Inga form a coupled system, a collective system? Inga certainly meets the criteria given by C and C.

1. Inga is readily available to Olaf and Olaf typically invokes Inga on a variety of daily details. ‘Inga, what time is my appointment with the Dean?’ ‘Inga what is the name of my teaching assistant?’
2. The information that Inga provides Olaf is more or less automatically endorsed. In fact, Olaf has come to so rely on Inga that he does even trust his biological memory. He often asks Inga to verify things that he has biologically recalled. ‘I think I have an appointment Thursday. Is this correct?’
3. Because I have stipulated that Inga is always with Olaf the information contained in Inga is easy for Olaf to access. Indeed, Inga is much more convenient and reliable than Otto’s notebook. After all, Otto needs to retrieve the notebook and then locate where he has put the address. He might forget to bring his notebook or it might go through the wash. This is not likely to happen with Inga. Because Inga is an active participant in the coupled system of which she is a member her presence is more reliable than a mere artifact. A loving and committed, cognitive partner, Inga is always there – through sickness, health, and memory loss.
4. Finally, the information that is contained in Inga is information that Olaf previously endorsed at some time or another. Inga is not making it up as she goes along. Olaf is partly responsible for the storage of this information. ‘Inga, will you remind me that I have an appointment on Thursday at 4?’

It seems, then, that if C and C are correct, the mind not only extends to encompass non-biological artifacts, forming systems that support cognition and dispositional attitudes like belief, but it also occasionally forms collective systems that support cognition and belief. But are C and C correct? In the next section I consider some of the most pressing objections to C and C’s view. In recent work Clark has offered some compelling responses to these objections. But even if these objections lead one to be skeptical about the possibility of solipsistic coupled systems, they should not lead to skepticism about collective coupled systems. Collective systems are surprisingly resilient in the face of these objections.

4. Objections

The argument for the extended mind as developed by Clark and Chalmers rests substantially on the notion of functional equivalence. Otto’s notebook forms a coupled system with Otto because his notebook is said to be

functionally equivalent to Otto's short term memory. Likewise, Inga and Olaf form a coupled system because the interaction between them is *functionally equivalent* to that found in biological memory (or some part of it). One might argue, however, that coupled systems of either the *solipsistic* or the *collective* sort are simply too different to support this claim. As one reviewer of this article put it, "the system (comprised of Olaf and Inga) appears fundamentally different from a biological brain, insofar as it includes modules not present in the latter – such as those devoted to communication and perception – which are by no means irrelevant for the prediction and explanation of its behavior".⁶

There seem to be two separate issues here. The first is whether the testimony of Olga and Otto's notebook play the same role or have the same function as that contributed by biological memory. The second is whether Olga and the notebook *carry out their function in the same way* that biological memory does.⁷ An artificial heart and a biological heart carry out the same function (pump blood) but arguably they do so in different ways. Now it is true that the coupled systems comprised of Otto and his notebook and Olaf and Olga may differ in the way that they carry out their function (because they have different modules, for instance) but it is not clear that this means that they are not *functionally equivalent* to a purely biological memory system.

Clark and Chalmers presuppose a common sense functionalist approach. Functional equivalence will be determined by the "causal dynamics" of the case, where the causal dynamics are judged by folk psychological standards. From our perspectives as interpreters of behavior, Otto will be guided by the information in his notebook in the same manner as he would be guided by the information in his short term memory. He will do the same sorts of things, say the same sorts of things, and form the same sorts of propositional attitudes on the basis of the information retrieved from his notebook. This is a loose sense of "causal dynamics" but it is clearly the sense which fuels much of our everyday practice of predicting and explaining people's behavior. I can predict with some accuracy that after Otto looks at his notebook he will head in the direction of the museum. Likewise, in the case of Olaf and Olga I can predict with great accuracy that Olaf will meet with his assistant on the basis of the information he retrieved from Olga. As Dennett points out (1978), this practice is extremely successful despite the fact that we know very little about brain processes and "modules". If one means by "causal

dynamics" the actual causal mechanisms and requires for functional equivalence that the function be performed by the same mechanism then accurate judgments of functional equivalency, even among and within uncoupled biological systems, will be rare indeed.

One might press the worry about functional equivalence by pointing out that the Otto scenario as described by C and C involves a notion of memory that is antiquated.⁸ Biological memory is not a filing cabinet that contains static information waiting to be retrieved. We now know that biological memory is active. It often constructs rather than merely recalls information and we know that our current goals, moods, beliefs, and so on influence what is retrieved from memory. But Otto's notebook is passive. It does not function in the same way that biological memory does. It does not contribute to any reorganization of experience, synthesis with other information, and so on. Thus, Otto's notebook cannot serve the role that biological role serves, and therefore fails to meet the functional criteria for being part of Otto's mind.

Clark has responded to this worry in the following way (2004b). First, although Otto's notebook is passive, C and C have not claimed that Otto's notebook is Otto's long term memory, or that considered alone, Otto's notebook would be a cognitive system. Rather, Otto's notebook is part of a complex system that involves Otto's brain and nervous systems and probably other features of his environment. The notebook plays a role in the cognitive system and its passivity makes it no less a candidate for playing such a role than the passivity of certain neurons in the brain. As Clark puts it,

True, that which is stored in Otto's notebook... won't participate in the ongoing underground reorganizations, interpolations, and creative mergers that characterize much of biological memory. But when called upon, its immediate contributions to Otto's behavior still fit the profile of stored belief. Information retrieved from the notebook will guide Otto's reasoning and behavior in the same way as information retrieved from biological memory. (2004b, p. 21)

I am rather convinced by this response but note that the objection loses its force when we consider my case of collective systems. Unlike Otto's notebook Inga is active. The addresses, names, numbers, dates and so on which Inga contributes are likely to be subject to the same sort of "reorganizations, interpolations, and creative mergers" found in Olaf's biological memory and this is precisely because the information drawn upon is being stored in Inga's biological memory.

But the active contribution of Inga might fuel skepticism about the plausibility of collective systems. If the processes

⁶ Personal correspondence via review process.

⁷ The reviewer mixes these two issues together in the following comment as well: "If the notebook does not perform the same operations carried out by the human memory module, how can it guide Otto's behavior in the same way as memory would". (personal correspondence) Could not it guide it in the same way but via different operations?

⁸ This objection appears in Clark (2004b) and is attributed to Terry Darnall.

are occurring “inside” Inga’s head then what we have is a mere transference of information. We have two systems here, Inga’s and Olaf’s. But imagine that the retrieval of information and the reorganization is done by Inga and Olaf jointly? Olaf cannot remember the last name; Inga cannot remember the first name. Olaf recalls *where* they met. Inga recalls *why* they met. Through a process of joint deliberation they jointly retrieve, via a process of joint reconstruction, the name of the person. The process of retrieval then is active and not found “inside” the heads of Inga and Olaf. Indeed, it is done via the joint activity of discussion and deliberation that occurs between Inga and Olaf.

What I have just described is not only possible but actual. It is called, in the literature on social cognition, transactive memory and has been well documented by Wegner (1987, 1995) and Wegner et al. (1985). A transactive memory system involves the operation of individual memory systems and the interaction between them. This system goes through the same stages that occur at the individual level: encoding, storage, and retrieval. Encoding at the collective level occurs when members discuss information and determine the location where it will be stored in the group and in what form it will be stored. Retrieval involves identification of the location of the information. Retrieval is transactive when the person who holds an item internally is not the person who is asked to retrieve it. Full retrieval may not occur until various sources are checked.

At this point it seems that merely asking someone else a question is enough to establish a collective memory system. But transactive memory is much more complex. First, individuals in a transactive memory system must know something about each other’s domains of expertise. They need to be aware of where information is stored and the storage capabilities of individuals in the group. Known experts in the group are held responsible for the encoding, storage, and retrieval of domain specific information. Other members contribute to the storage of information by directing new information to the appropriate expert. When there are no clear experts other ways of assigning responsibility for the information are used. The person who first introduces the information may, by default, be held responsible for encoding, storing and retrieving the new information.

And although Wegner does not say this explicitly, transactive memory systems are formed through a process of interaction over time. If I ask a policeman for directions, we have not formed a transactive memory even though I have retrieved information from him. Individuals must participate in the encoding, storage, and retrieval of information. They must be involved in the allocation of information to specific experts, for instance, and in the determination of what information will be stored. Further, transactive memory arises in a group when they are engaged in a common goal or share a common perspective; the purpose being to share cognitive

responsibility and work. As a rule we can say that transactive memory systems are properties of groups. Since the policeman and I do not form a group it is not surprising that we do not form a transactive memory system.

Wegner’s point is that people often become epistemically dependent on others and this interdependence often forms a ‘knowledge-holding’ system that is larger and more complex than either of the individuals’ own memory. Indeed, group retrieval of information appears to be a more reliable form of memory than individual memory. Transactive memory, then, provides us with an actual instance of a collective system and it puts us in a position to consider other objections that have been raised to the extended mind hypothesis.

Adams and Aizawa (2001) have argued that a computer or some other artifact could not constitute part of the human mind because the symbols in the human mind have intrinsic content whereas Otto’s notebook has only derived content. Echoing Searlean views about content Adams and Aizawa write:

The representational capacity of orthography is in this was derived from the representational capacities of cognitive agents. By contrast, the cognitive states in normal cognitive agents do not derive their meaning from conventions or social practices. . . (2001, p. 48)

Human cognition according to this line involves a special kind of content. Since computers can only have derived content they cannot be part of human cognition. Clark responds to this objection by pointing out that even if there is a sense in which human cognitive states enjoy some sort of “intrinsic” content, the issue that needs to be established is why everything that is to count as an individual cognitive system must be composed solely of states of affairs that have intrinsic content. Clark provides the following example:

...suppose we are busy (as part of some problem solving routine) imagining a set of Venn Diagrams/Euler Circles in our mind’s eye? Surely the set-theoretic meaning of the overlaps between say, two intersecting Euler circles is a matter of convention. Yet this image can clearly feature as part of a genuinely cognitive process. (2004b, p. 10)

I find this line of reasoning to be a compelling response to Adams and Aizawa. But one need not have a view about these matters in order to see that the objection does not seem to have any force against collective systems. If there is such a thing as “intrinsic” content, then collective systems can have it in virtue of the fact that collective systems are constituted, in part, by individual brains and the relations between them. Otto’s notebook has only derived content. Olga, however, has intrinsic content and the system comprised of Olga and Otto would allow for the supervenience of dispositional states that have intrinsic content.

There have been several other objections to the extended mind hypothesis that center around the notions of responsibility and control. Keith Bulter raises the following worry:

...there can be no question that the locus of computational and cognitive control resides inside the head of the subject and involves internal processes in a way quite distinct from the way external processes are involved. If this feature is indeed the mark of a truly cognitive system, then it is a mark by mean of which the external processes Clark and Chalmers point to can be excluded. (Butler, 1998, p. 205)

Butler suggests that control is the distinguishing mark of the mental. Because computers, palm pilots, and pencils, are not the locus of control they are not part of the mind. Again C and C have a compelling response to this concern. If we apply the locus of control criterion inside the head we end up shrinking the mind almost to the point of disappearance. Do we now discount any neural subsystem as part of the mind because that subsystem is not the ultimate arbiters of action and choice? 'Suppose a small part of my frontal lobe has "the final say" does this mean that the mind is to be identified only with my frontal lobe? What if Dan Dennett and others are right and there is no subsystem that has the final say – has the mind just disappeared?' (Clark, 2004b, p. 24)

The control worry seems less pressing when we consider the phenomenon of transactive memory. Control is distributed among agents in transactive memory systems. In a very real sense the group, rather than any one individual in the group, controls the information retrieval process. Distributed control is also present in cases of joint deliberation and research. Consider research teams. There is often a distribution of labor in these cases and people are allocated certain cognitive tasks. When they engage in joint decision-making and problem solving there may be no one person who has the final say. The consensus they achieve is a joint effort – it is the group's final say. The locus of control in some cases is not to be identified with the first person singular perspective but with the first person plural perspective.⁹ What seems to be motivating Butler's worry that if we acknowledge that the mind extends to the environment we relinquish control to artificial objects like computers; artifacts that lack agency. But this is not a possibility in the case of collective systems. Control is distributed among many agents.

The shift of control in collective systems allows us to respond to yet another criticism raised to the extended mind hypothesis. Several philosophers have raised the

worry that the extend mind hypothesis commits one to a sort of no-self theory or at least a bizarre, counterintuitive, view of the self.¹⁰ Further, the idea of responsible agency seems to become lost once one admits that the mind can extend and encompass aspects of the environment. Dartnall puts the worry this way: "If I dig a hole in my garden with my spade... my spade and I do not get the prize 'for the best hole in the garden'. I get the prize, even though I could not have done the digging without the spade" (found in Clark, 2004b, p. 7). For Dartnall, computers and other cognitive prosthetics are merely tools used by agents.

Clark responds by considering whether we would conceive of the biological arm and hand as a tool? The spade example is a fitting counterexample but only because we do not use the spade on a regular basis and it is not always readily available. But suppose I win the best hole in the garden prize by digging with my prosthetic hand. True, my hand and I do not get the prize – I get the prize; the hand is part of me. If the spade were a constant in my life the way prosthetics are to those that have them, our intuitions may change about what makes something part of one's self. The intuition fueling Dartnall's comments might be that the real self is "inside" the head. Here's what Clark has to say about this:

Go into the head in search of the real self and you risk cutting the cognitive cake ever thinner, until the self vanishes from your grasp. For there is no single circuit in there that makes the decisions, that does the knowing, or that is in any clear sense the seat of the self. At any given moment, lots of neural circuits (but not all) are in play. The mix varies across time and task, as does the mix between bodily and neural activity and all those profoundly participant non-biological props and aids. (2004b, p. 15)

Again, I am rather convinced by this response. The idea that my body is a mere tool is more counterintuitive to me than the idea that a spade could become part of me. But notice that objections related to responsible agency and self seem less pressing in the case of collective systems.

Indeed, we often hold groups morally and epistemically responsible. An alteration of Dartnall's digging example makes the case quite clearly. Suppose it is not a spade but another person digging the hole with me. We cooperate, coordinating our actions and so on. It seems perfectly reasonable in this case that *we* should receive the prize for the best hole in the garden. Likewise, groups are often held epistemically responsible for their joint cognitive endeavors. When colleagues write a joint paper they are honored and recognized for their joint achievement. There seems to

⁹ See Tollefsen (2004) for a similar line of reasoning.

¹⁰ In particular, Dartnall (2004).

be no reason not to recognize collective systems as capable of responsible agency.

But what does this mean for our conception of the self? Where is the self-located? Is the notebook part of Otto's self? And, to apply the worry to collective systems, is Inga part of Olaf? Is Olaf part of Inga? Where does Inga begin and Olaf end? Is there a collective person or self, (call it Olga), present?

It is here that I have to admit that things get tricky. Clark responds to these worries by advocating the position that our common sense ideas of person, self, and agency are forensic notions. That is, they are concepts whose application is matter of pragmatic convenience rather than metaphysical necessity. I would like to resist instrumentalism with respect to personhood and self. Luckily, one does not need to embrace it in order to avoid these worries. It is quite possible that there is more to personhood and selfhood than cognition and intentionality – consciousness, for instance.¹¹ If this is so then the fact that the mind can sometimes extend beyond the skin to form coupled systems (solipsistic or collective) does not necessarily mean that the self is also outside the skin or that there are collective persons.

5. Collective minds, collective systems, and explanatory power

The title of my paper, recall, is “From extended mind to collective mind” but I have restricted my discussion to collective systems. So, are collective systems, collective minds? C and C move from extended to cognition to extended mind by arguing that Otto's dispositional beliefs are determined, in part, by the contents of his notebook. If the vehicle of both cognition and belief is to be found, at times, outside of the head, then the mind is, at times, outside of the head. The mind is where cognition and

belief are. If this is so, then, yes, there are collective minds.¹²

But this raises some additional questions. Is Inga merely a part of Olaf's mind? Is Olaf part of Inga's mind? Where do the individual minds end and the collective mind begin? I think these questions are motivated by a certain view of the mind that has been with us for far too long – call it, substantialism. According to substantialism the mind is a thing or substance – an *object* of study. In its contemporary form substantialism does not imply that the mind is immaterial, rather it is simply the view that the mind could be the sort of thing contained by the skin or composed of parts. To ask “is Inga part of Olaf's” mind is to think of the mind as a substance – a whole, with parts. To ask “whose mind is it?” is to presuppose that the mind is an object that can be owned. I am inclined to agree with Ryle (1949) that substantialism makes a serious category error. “Mind” is not a name for a substance; rather it names a whole host of cognitive processes, dispositional states, connotative and agential behavioral dispositions. The picture of mind that comes out of C and C's work is that some of these states and processes supervene on features outside of the body. I have argued in this paper that this will include, in some cases, other agents. If the mind is a collection of processes and states then the divisions between minds or the merging of minds will be determined according to the role agents and artifacts play in these processes and will be fueled to a great extent by explanatory needs.

But is there any explanatory need for positing collective minds or collective systems? Although the explanatory issue deserves a paper of its own, let me briefly respond to one of the few sustained attacks on the collective mind hypothesis. This attack specifically targets the issue of explanatory power. In *Boundaries of the Mind* (2004) Robert Wilson argues that appeals to a group or collective mind are often confused attempts to state a version of what

¹¹ There are some who would argue that consciousness is the essence of mind and thus, a being that lacks consciousness lacks mind. As one reviewer of this paper put it: “The issue is not whether the thesis that interpreting a collection of individuals and their actions and tools as if they constituted one distributed mind has explanatory power, but whether it is true that they constitute a mind”. The truth of that is dependent on one thing and one thing only: Can the collective entity feel? If it cannot, it is not a mind” (personal correspondence). It is not clear to me that one can settle what a mind is by a priori means or by reflecting on our experience as human minds. It is clear that human minds “feel” (at least some of the time) and that the zombies of philosophical thought do not. But it is not clear that minds without phenomenological experience are conceptually impossible. The fact that human minds feel and collective systems (and zombies) do not should not lead us to skepticism about the possibility of collective minds. Rather, we should conclude, as we do in the case of animal minds, that there are different sorts of minds. If one wants to insist on using the word “mind” to mean only those things that “feel”, then I would agree that collective systems are not minds in that limited sense. But I do not believe there is a non-question begging argument to be had for using the word in this restricted sense.

¹² One might object that the move from collective systems to collective minds is too quick. Something could be a cognitive system, a computer for instance, but fail to “have” a mind or “have” mental states. According to this line of thought, there is a distinction between *being* a cognitive system (or a collective system) and *having* a mind or mental states. The latter requires something more. What more is needed? I think the distinction between being a cognitive system and having mental states does not mark an ontological difference but a normative one. We often talk of “having” mental states – “I have a belief, I have a desire. . .” What this talk reveals is the normative nature of folk psychology. We take ownership of our thoughts, feelings, and ideas and in many cases we are held accountable for these states both morally and epistemically. To “have” mental states is to be the locus of authority and responsibility. If there are cognitive systems that do not “possess” mental states it is because they are not subject to certain kinds of appraisal and this is because they do not engage in the sorts of interpersonal relations that give rise to such appraisals. It seems clear that we do hold groups epistemically and morally responsible and that we often assess the consistency and coherency of group decisions and attitudes. Unlike solipsistic coupled systems, the notion of responsibility is readily applied to collective coupled systems (see Tollefsen, 2003, 2004, for a more extended discussion of the moral and epistemic responsibility of groups).

Wilson calls, the *social manifestation thesis*. The social manifestation thesis refers to the claim that individuals have properties which are manifest only when those individuals form part of a group of a certain type. Wilson contrasts this thesis with the collective mind thesis which is the claim that groups have properties, including mental states, which are not reducible to the individual states of individuals. Wilson argues that appeals to group minds at the turn of the twentieth century often confuse the two theses and he suggests that, for instance, the work of Gustav LeBon in *The Crowd* (1895) provides support only for the social manifestation thesis and that the social manifestation thesis is sufficient for explaining the behavior of crowds. Since my interest is in contemporary appeals to collective minds I will not address Wilson's criticisms of the collective mind tradition.

Having distinguished between the group mind hypothesis and the social manifestation hypothesis and how these two have been confused by early thinkers at the turn of the 20th century, Wilson directs his attention to two contemporary appeals to group minds; one found in the work of David Sloan Wilson (2002, 1997) and the other in the work of Douglas (1986). I consider his criticisms of Sloan Wilson first.

Sloan Wilson has recently advocated the group mind hypothesis in the biological sciences. His appeal to group minds is in keeping with his emphasis on group level adaptations and his theory of group selection. Sloan Wilson extends his notion of group-level adaptation to cognitive phenotypes such as cognition and decision-making. He writes,

Group-level adaptations are usually studied in the context of physical activities such as resource utilization, predator defense, and so on. However, groups can also evolve into adaptive units with respect to cognitive activities such as decision making, memory, and learning. As one example, decision making is a process that involves identifying a problem, imagining a number of alternative solutions, evaluating the alternatives, and making the final decision on how to behave. Each of these activities can be performed by an individual as a self-contained cognitive unit but might be performed even better by groups of individuals interacting in a coordinated fashion. At the extreme, the groups might become so integrated and the contribution of any single member might become so partial that the group could literally be said to have a mind in a way that individuals do... (1997, p. 131)

Sloan Wilson has also suggested that religion is a group level adaptation.

Robert Wilson argues, however, that in order for religion and group decision making to provide a defense of the group mind hypothesis Sloan Wilson needs to do two things. First, he must show that it is not just religious groups but religious ideas that are adaptations. This would show that there are group-level mental properties surviving

the process of selection. In the case of group decision making he must show that there are mental or cognitive properties that are adaptations. Second, these mental properties must be properties of groups and not simply the individuals within the group.

These seem to be legitimate requirements, but when Wilson turns his focus to the issue of group decision making he adds an additional constraint. Wilson writes, "With respect to human decision making, he (Sloan Wilson) would seemingly need to show that this functions at the group level by individuals relinquishing their own decision-making activities. For it is only by doing so that he could point to a group-level psychological characteristic that is, in the relevant sense, emergent from individual-level activity". (2004, p. 297)

It is not clear why this constraint is introduced. He appeals to the fact that historically the collective mind hypothesis has been linked to the emergentist tradition. Emergentism is the view that, at a higher level of organization, properties emerge which are unique and are not to be found at the lower level.

We have already seen that both the collective psychology and superorganism tradition has an emergentist view of the nature of groups (and thus group minds): groups are more than the sum of their individual parts, and having a group mind is more than having a group of individuals with mind. (2004, p. 297)

But collective minds on Sloan Wilson's account are more than simply a collection of individuals with minds; just as an individual mind is more than a collection of neurons. Individuals in a decision making group interact and coordinate with each other. The processes that take place between individuals will be processes of the group. Further, it does not follow that because early accounts of collective mind developed from emergentist views that contemporary accounts of collective mind do so. Indeed, most discussions of collective minds do not see collective mental properties and mental states as emergent, but as supervenient on a set of individual mental states and processes or as realized by a set of individual mental states and processes. Emergentism has fallen out of favor as a theory of individual mental properties and, thus, it is no longer the model for understanding the relationship between individual and group level processes and properties. Therefore, there is no reason to require that Sloan Wilson's account of group cognition and collective (religious) ideas involve the relinquishing of individual level decision making processes or individual religious beliefs.

There is a further point to be made in defense of the appeal to group cognition in the work of Sloan Wilson. Sloan Wilson is concerned with identifying certain adaptive cognitive *processes* that occur at the group level. But Robert Wilson seems to miss this when he focuses on group decisions rather than group decision making processes. For instance, Wilson provides an example of a club that makes a decision via majority vote and insists that this

could be explained via appeal to individuals using their own cognitive processes to arrive at an opinion and then voting on that opinion. But decisions, conceived of as static states, are not what Sloan Wilson has in mind when he talks about group level decision making. He is talking about the deliberative process, the giving and taking of reasons, which literally takes place outside the minds of individuals and in the realm of public discourse. This process often contributes to the formation of individual opinion but it also forms consensus opinions which may or may not be reducible to a majority. In his criticisms of Sloan Wilson, Wilson fails to make a distinction between a decision which is the end of result of a decision process and the cognitive process of deliberation which is exhibited in the complex interactions of individuals. It is the cognitive process which is adaptive and it is this process that can be understood as supervening on, rather than emerging from, the interaction of individuals.

I have not tried to show here that Sloan Wilson's appeal to collective cognition has more explanatory power than competing accounts. I have simply tried to address Wilson's criticisms. Whether Sloan Wilson's theory is explanatorily powerful is an empirical question, one that will have to be answered after we see the theory developed and applied. Sloan Wilson's work is conjectural. He is introducing a new way of understanding evolutionary processes. The computational theory of mind was, at one time, a conjectural hypothesis as well and it took time to reveal its explanatory power or, in some cases, its lack of. Given the problems I have raised with Wilson's attack on the group mind hypothesis in contemporary biology, and the fact that this theory is in its infancy, dismissal of it seems premature.

I turn now to Wilson's criticisms of Mary Douglas's *How Institutions Think* (1986). Here too we see a premature dismissal, not of Douglas's work but, of the explanatory role of the collective mind hypothesis in the social sciences. Douglas, herself, appeals to the notion of a group mind in a tentative fashion and Wilson may be correct that her account of institutions really is an expression of the social manifestation thesis but Douglas's work, however, is almost 20 years old and fails to exhibit the sort of sophisticated work that social scientists are now doing with the collective mind hypothesis. Both Wegner's work and Edwin Hutchins' work, not to mention a great deal of research in organizational theory, involves the application of cognitive models such as connectionism to groups in order to understand higher level social phenomena. The dismissal of Douglas' work, as it is not representative of the current work in the social science, is no indication that the group mind hypothesis lacks explanatory power. Hutchins' (1991, 1995) work on navigational teams suggests that viewing the team itself as a cognitive system provides a richer explanation of the cultural and cognitive dynamics present on naval vessels than one that focuses merely on individual cognition. The work of Giere and Moffatt (2003) and Thagard (1993) suggests that large scale

scientific research should be viewed as being undertaken by cognitive systems comprised of many agents and many artifacts and that this provides a richer understanding of scientific research than an individualistic and atomistic approach to science. Of course, to cite the highly developed use of the collective mind hypothesis in current social science research does not provide a defense of its explanatory power. These projects will have to be judged on their own merits. But to date there has been no sustained attack on this explanatory framework such that it would lead us to dismiss the hypothesis altogether. Wilson's limited discussion of the work of Sloan Wilson and Douglas certainly does not support the claim that there is no explanatory need to posit a collective mind. The jury, as they say, is still out.

6. Conclusion

In this paper, I have tried to remove one of the motivations for rejecting the idea that groups could be the legitimate bearers of mental states. If the mind and its processes are not bounded by the skin then this opens up the possibility that groups could themselves form systems that can sustain cognitive properties and processes. The thesis of the paper, however, is somewhat limited. I have only argued for the possibility and plausibility of collective minds. Further, although I have attempted to defend the hypothesis against the charge of explanatory weakness, I have not definitively established its explanatory power.¹³ What I have suggested however is that, like all theories of cognition, explanatory power is something that can be judged only as a theory develops and is applied. The collec-

¹³ One might argue that we can rule out the explanatory power of collective minds right off by citing the success of explanations using non-mentalistic and non-intentional terms. The suggestion is that we ought to take the "design stance" (Dennett, 1978) toward collective cognitive systems or even the "physical stance". No doubt the design stance and physical stance have proved explanatorily fruitful in understanding certain aspects of collective systems. These explanatory approaches have been useful in understanding individual cognitive systems as well. What stance one takes depends on one's explanatory needs. The "intentional stance" meets a need for general explanation that cannot be found at these lower levels of description. We might use the design stance or the physical stance to understand the inner workings of a particular collective system but given the possibility of multiple realizability we need a more general description of these collective processes. If, for instance, we want to talk about collective systems within science (research teams, for instance) we may want to make generalizations across various groups. Folk psychological concepts provide the requisite level of generality. Do we need the concept "mind"? As I have suggested this concept marks, not a substance, but a collection of processes – memory, deliberation, practical reasoning, and so on. I do not think cognitive scientists find the term "mind" all that useful but they continue to use folk psychological concepts. My point here is that just as there is an explanatory need for the intentional stance when explaining individual action, there is a need for the intentional stance in explaining collective action (see Tollefsen, 2002, for a discussion of the explanatory power of collective mental states in the social sciences). Of course, the explanatory power of folk psychology has been and continues to be challenged, but this is not the place to engage in this debate.

tive mind hypothesis is in its infancy. Explanatory power is something to be judged only as the theory matures. I hope to have established on theoretical grounds that it is a hypothesis that ought to be allowed to mature.

Acknowledgements

Work on this paper was supported by a Faculty Research Grant from the University of Memphis. I thank the anonymous reviewers for their insightful and challenging criticisms, the guest editors for their guidance and patience, and the participants of Collective Intentionality IV for their helpful comments and questions.

References

- Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14, 43–64.
- Bratman, M. (1993). Shared intention. *Ethics*, 104, 97–113.
- Butler, K. (1998). *Internal affairs: a critique of externalism in the philosophy of mind*. Dordrecht, The Netherlands: Kluwer.
- Burge, T. (1979). Individualism and the mental. In French, Uehling, & Wettstein (Eds.), *Midwest studies in philosophy* (Vol. IV, pp. 73–121). Minneapolis: University of Minnesota Press.
- Chalmers, D., & Clark, A. (1998). The extended mind. *Analysis*, 58, 10–23, Reprinted in Chalmers, 2002, 643–652.
- Chalmers, D. (2002). *Philosophy of mind: classical and contemporary readings*. Oxford, UK: Oxford University Press.
- Clark, A. (2001a). *Being there*. Cambridge, MA: MIT Press.
- Clark, A. (2001b). Reasons, robots, and the extended mind. *Mind and Language*, 16, 121–145.
- Clark, A. (2003). *Natural born cyborgs*. Oxford, UK: Oxford University Press.
- Clark, A. (2004a). Memento's revenge: the extended mind revisited. In R. Menary (Ed.), *The extended mind*. Amsterdam, The Netherlands: John Benjamins.
- Clark, A. (2004b). "Author's reply" to symposium on Natural-born cyborgs. *Metascience*, 13(2).
- Dennett, D. (1978). *Intentional stance*. Cambridge, MA: MIT Press.
- Dartnall, T. H. (2004). We have always been... cyborgs. *Metascience*, 13(2), 139–148.
- Douglas, M. (1986). *How institutions think?*. Syracuse NY: Syracuse University Press.
- Giere, R., & Moffatt, B. (2003). Distributed cognition: where the cognitive and social merge. *Social Studies of Science*, 33(2), 1–10.
- Gilbert, M. (2002). Belief and acceptance as features of groups. *Protosociology*, 16, 35–69.
- Gilbert, M. (2003). The structure of the social atom: joint commitment and the foundation of human social behavior. In F. Schmitt (Ed.), *Socializing metaphysics* (pp. 39–64). Maryland: Rowman and Littlefield.
- Hutchins, E. (1991). The social organization of distributed cognition. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 283–307). Washington, DC: American Psychological Association.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.
- LeBon, G. (1895). *The crowd: A study of the popular mind* (2nd ed.). Dunwoody, GA: Norman S. Berg, 1986.
- McDowell, J. (1992). Putnam on mind and meaning. *Philosophical Topics*, 20(1), 35–48.
- Putnam, H. (1975). The meaning of 'meaning' Language, mind and knowledge. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science* (Vol. 7, pp. 131–193). Minneapolis: University of Minnesota Press.
- Ryle, G. (1949). *The concept of mind*. Chicago: The University of Chicago Press.
- Searle, J. (1990). Collective intentions and actions. In P. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 401–415). Cambridge, MA: Bradford Books, MIT press.
- Searle, J. (1995). *The construction of social reality*. New York, NY: Free Press.
- Thagard, P. (1993). Societies of minds: science as distributed computing. *Studies in History and Philosophy of Science*, 24, 49–67.
- Tollefsen, D. (2002). Collective intentionality and the social sciences. *Philosophy of the Social Sciences*, 32(1), 25–50.
- Tollefsen, D. (2003). Participant reactive attitudes and collective responsibility. *Philosophical Explorations*, 6(3), 218–234.
- Tollefsen, D. (2004). Collective epistemic agency. *Southwest Philosophy Review*, 20(1), 55–66.
- Tuomela, R. (1992). Group beliefs. *Synthese*, 91, 285–318.
- Wegner, D. M. (1987). Transactive memory: a contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–208). New York: Springer.
- Wegner, D. M. (1995). A computer network model of human transactive memory. *Social Cognition*, 13, 319–339.
- Wegner, D. M., Giuliano, T., & Hertel, P. (1985). Cognitive interdependence in close relationships. In W. J. Ickes (Ed.), *Compatible and incompatible relationships* (pp. 253–276). New York: Springer.
- Wilson, D. S. (1997). Incorporating group selection into the adaptationist program: a case study involving human decision making. In J. Simpson & D. Kendrick (Eds.), *Evolutionary social psychology*. Hillsdale, NJ: Erlbaum.
- Wilson, D. S. (2002). *Darwin's Cathedral: Evolution, religion, and the nature of society*. Chicago: University of Chicago Press.
- Wilson, R. (2004). *Boundaries of the mind*. Cambridge, UK: Cambridge University Press.