The Extended Mind

Andy Clark & David Chalmers

Department of Philosophy

Washington University

St. Louis, Missouri 63130

andy@twinearth.wustl.edu

dave@twinearth.wustl.edu

April 10, 1995

1 Introduction

Where does the mind stop and the rest of the world begin? The question invites two standard replies. Some accept the intuitive demarcations of skin and skull, and say that what is outside the body is outside the mind. Others are impressed by the arguments of Putnam and Burge that the truth-conditions of our thoughts "just ain't in the head",[1] and hold that this externalism about meaning carries over into an externalism about mind. We propose to pursue a third position. We will advocate an externalism about mind, but one that is in no way grounded in the debatable role of truth-conditions and reference in fixing the contents of our mental states. Rather, we advocate an *active externalism*, based on the active role of the environment in driving cognitive processes.

[1] H. Putnam, "The Meaning of 'Meaning"', in *Language, Mind, and Knowledge*, ed. K. Gunderson (Minneapolis: University of Minnesota Press, 1975); T. Burge, "Individualism and the Mental", *Midwest Studies in Psychology* 4 (1979):73-122.

2 Extended Cognition

Consider three cases of human problem-solving:

(1) A person sits in front of a computer screen which displays images of various two-dimensional geometric shapes and is asked to answer questions concerning the potential fit of such shapes into depicted "sockets". To assess fit, the person must mentally rotate the shapes to align them with the sockets.

(2) A person sits in front of a similar computer screen, but this time can choose either to physically rotate the image on the screen, by pressing a rotate button, or to mentally rotate the image as before. We can also suppose, not unrealistically, that some speed advantage accrues to the physical rotation operation.

(3) Sometime in the cyberpunk future, a person sits in front of a similar computer screen. This agent, however, has the benefit of a neural implant which can perform the rotation operation as fast as the computer in the previous example. The agent must still choose which internal resource to use (the implant or the good old fashioned mental rotation), as each resource makes different demands on attention and other concurrent brain activity.

Cases could have been multiplied. We might have imagined genetically engineered versions of case (3), or human mutants who spontaneously exhibited such faster mental rotation skills. Such minor variations may help those who set store by some notion of the 'natural' array of cognitive capacities found in human brains. Either way, the question remains: what are we to say about the various cases? While case (2) allows for computations to be distributed across agent and computer, case (3) displays (by hypothesis) the same sort of computational structure, but this time internalized within the agent. If the latter case is (as it seems to be) a case of cognitive processing, by what right do we count case (2), using the external rotation button, as fundamentally different? We cannot simply point to the skin/skull boundary as justification, since the legitimacy of that boundary is precisely what is at issue. But nothing else seems different.

The kind of case just described is by no means as exotic as it may at first appear. It is not just the presence of advanced external computing resources which raises the issue, but rather the general tendency of human reasoners to lean heavily on environmental supports. Thus consider the use of pen and paper to perform long multiplication,[2] the use of physical re-arrangements of letter tiles to prompt word recall in Scrabble,[3] the use of instruments such as the nautical slide rule,[4] and the general paraphernalia of language, books, diagrams, and culture. In all these cases the individual brain performs some operations, while others are delegated to manipulations of external media. Had our brains been different, this distribution of tasks would doubtless have varied.

In fact, even the mental rotation cases described in scenarios (1) and (2) are real. The cases reflect options available to players of the computer game Tetris. In Tetris, geometric shapes descend from the top of the screen and must be built into compact walls as the game progresses. A falling shape must be directed, rather rapidly, to an appropriate slot in the emerging structure. The physical rotation operation is available to players. David Kirsh and Paul Maglio[5] calculate that the physical rotation of a shape through 90 degrees takes about 100 milliseconds, plus about 200 milliseconds to select the rotate button. To achieve the same result by mental rotation takes about 1000 milliseconds. Kirsh and Maglio go on to present compelling evidence that this physical rotation operation is used not just to position a shape ready to fit a slot, but often to help *determine* whether the shape and the slot are compatible. The latter use constitutes a case of what Kirsh and Maglio call an 'epistemic action'. *Epistemic* actions alter the world so as to aid and augment cognitive processes such as recognition and search. Merely *pragmatic* actions, by contrast, alter the world because some physical change is desirable for its own sake (e.g.,

[2] J.L. McClelland, D.E. Rumelhart, and G.E. Hinton, "The appeal of parallel distributed processing", in *Parallel Distributed Processing, Volume 2*, ed. McClelland and Rumelhart (Cambridge, MA: MIT Press, 1986); A. Clark, *Microcognition*, (Cambridge, MA: MIT Press, 1989), Chapter 6.

[3] D. Kirsh, "The Intelligent Use of Space", *Artificial Intelligence* (to appear).

[4] E. Hutchins, *Cognition in the Wild* (Cambridge, MA: MIT Press, 1995).

putting cement into a hole in a dam).

Epistemic action, we suggest, demands spread of *epistemic credit*. If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process. Cognitive processes ain't (all) in the head!

3 Active Externalism

The central feature of these cases is that the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. All the components in such a system play an active causal role, and together they govern behavior in the same sort of way that cognition usually does. Remove the external component and the system's behavioral competence will drop, just as it would if we removed part of its brain. Our thesis is that this sort of coupled process counts equally well as a cognitive process, whether or not it is wholly in the head.

Notice how different this sort of externalism is from the variety advocated by Putnam and Burge. When I believe that water is wet and my twin believes that twin water is wet, the external features responsible for the difference in our beliefs are distal and historical, at the other end of a lengthy causal chain. Features of the *present* are not relevant at all— as Burge points out, if I happen to be surrounded by XYZ right now (maybe I have been suddenly transported to Twin Earth), my beliefs will still concern standard water, because of my history. In these cases, the relevant external features are *passive*—because of their distal nature, they play no role in driving the cognitive process in the here-and-now. This is reflected by the fact that the actions performed by me and my twin are physically indistinguishable.

In the cases we describe, by contrast, the relevant external features are *active*, playing a

[5] D. Kirsh and P. Maglio, "On Distinguishing Epistemic from Pragmatic Action", *Cognitive Science*, 18, 513-49.

crucial role in the here-and-now. Because they are coupled with the human organism, they have a direct impact on the organism and on its behavior. In these cases, the relevant parts of the world are *in the loop*, not dangling at the other end of a long causal chain. Concentrating on this sort of coupling leads us to an *active externalism*, as opposed to the *passive externalism* of Putnam and Burge.

Many have complained that even if Putnam and Burge are right about the externality of truth-conditional content, it is not at all clear that these external aspects play a causal or explanatory role in the generation of action. In counterfactual cases where internal structure is held constant but these external features are changed, behavior looks just the same; so internal structure seems to be doing the crucial work. There is no danger of active externalism falling into a similar problem. The external features in a coupled system play an ineliminable role—if we retain internal structure but change the external features, behavior may change completely. The external features here are just as causally relevant as typical internal features of the brain.[6]

By embracing an active externalism, we allow a much more natural explanation of all sorts of actions. It is natural to explain my choice of words on the Scrabble board, for example, as the outcome of an extended cognitive process that centrally involves the rearrangement of tiles on my tray. Of course, one could always try to explain my action in terms of internal processes and a long series of "inputs" and "actions" characterizing the interaction with an external object, but this explanation would be needless complex. If an isomorphic process were going on in the head, we would feel no urge to characterize it in this cumbersome way.[7] In a very real sense, the re-arrangement of tiles on the tray is not part of action; it is part of *thought*.

---

[6]It may be that much of the appeal of externalism in the philosophy of mind stems from the intuitive appeal of active externalism. The analogies that are often made in support of externalism often involve the explanatory relevance of external features in a coupled system. And there is considerable appeal in the idea that the boundaries at skull and skin are arbitrary. But these intuitions sit uneasily with the letter of the externalism of Putnam and Burge. Upon examination, in these cases the immediate environment is irrelevant, and all that really counts is history. Debate has focused on the question of whether mind must be in the head, but the more relevant question in assessing these examples might have been, must mind be in the present?

[7]Interestingly, Herbert Simon once suggested that we view internal memory as, in effect, an external resource upon which "real" inner processes operate. "Search in memory," he comments, "is not very different from search

Indeed, the view we advocate here is reflected by a growing body of research in cognitive science. In areas as diverse as the theory of situated cognition,[8] studies of real-world-robotics,[9] dynamical approaches to child development,[10] and research on the cognitive properties of collectives of agents,[11] cognition is often taken to be continuous with processes in the environment. Thus, in seeing cognition as extended one is not merely making a terminological decision; in these cases, it makes a significant difference to the methodology of scientific investigation. In effect, explanatory methods that might once have been thought appropriate only for the analysis of "inner" processes are now being adapted for the study of the outer, and there is promise that our understanding of cognition will become richer for it.

Some may find this sort of externalism unpalatable. One reason may be that many identify the cognitive with the conscious, and it seems far from plausible that consciousness extends outside the head in these cases. But not everything that occurs in the brain, and which constitutes a cognitive process in standard usage, is tied up with conscious processing. It is widely accepted that all sorts of processes that beyond the borders of consciousness nevertheless play a crucial role in cognitive processing: in the retrieval of memories, linguistic processes, and skill acquisition, for example.[12] So the mere fact that external processes are external where consciousness is internal is no reason to deny that those processes are cognitive.

More interestingly, one might argue that what keeps real cognition processes in the head is the requirement that cognitive processes be *portable*. Here, we are moved by a vision of what

---

of the external environment. Simon's view at least has the virtue of treating internal and external processing with the parity they deserve, but we suspect that on his view the mind will shrink too small for most people's tastes. See H. Simon, *The Sciences of the Artificial* (Cambridge, MA: MIT Press, 1981), and also the exchange between Stefik and Simon in *Contemplating Minds*, eds. W. Clancey, S. Smoliar, and M. Stefik (Cambridge, MA: MIT Press, 1994), pp. 239-242.

[8]See e.g. L. Suchman, *Plans and Situated Actions* (Cambridge, UK: Cambridge University Press, 1987).

[9]E.g., R. Beer, *Intelligence as Adaptive Behavior* (New York: Academic Press, 1989).

[10]E. Thelen and L. Smith, *A Dynamic Systems Aproach to the Development of Cognition and Action* (Cambridge, MA: MIT Press, 1994).

[11]E. Hutchins, *Cognition in the Wild*, (Cambridge, MA: MIT Press, 1995.

[12]See e.g. E. Kandel, J. Schwarz, and T. Jessell, eds., *Principles of Neural Science*, 3rd edition (New York: Elsevier, 1991).

might be called the Naked Mind: a vision of the resources and operations we can always bring to bear on a cognitive task, regardless of whatever further opportunities the local environment may or may not afford us. The trouble with coupled systems, then, is that they are too easily *decoupled*. The true cognitive processes are those that lie at the constant core of the system; anything else is an add-on extra.

We are quite sympathetic to this objection. It seems clear that the brain (or perhaps, in this view, the brain and the body) constitutes a proper and distinct object of study and interest. And what makes it such is precisely the fact that it comprises some such set of basic, portable, cognitive resources. These resources may incorporate bodily actions as integral parts of some cognitive processes, as when we use our fingers to off-load working memory in the context of a tricky calculation, but they will not encompass the more contingent aspects of our external environment; the ones that come and go, like a pocket calculator.

Still, mere contingency of coupling does not seem enough to deny a process cognitive status. Imagine that in the future, we are able to plug various modules into our brain to help us out: a special module for additional short-term memory when we need it, for example, or a module to help with geometric reasoning. When a module is plugged in, the processes involving it are just as cognitive as if they had been there all along. Perhaps there is a sense in which they are extrinsic to the core system, but they are still part of thought.[13]

In any case, even if one were to make the portability criterion pivotal, active externalism would not be undermined. We have already seen that counting on our fingers has been let in the door, and it is easy to push things further. Think of the old image of the engineer with a slide rule hanging from his belt wherever he goes. What if people always carried a pocket calculator,

[13] Or consider the following passage from a recent science fiction novel: "I am taken to the system's department where I am attuned to the system. All I do is jack in and then a technician instructs the system to attune and it does. I jack out and query the time. 10:52. The information pops up. Always before I could only access information when I was jacked in, it gave me a sense that I knew what I thought and what the system told me, but now, how do I know what is system and what is Zhang?" M. McHugh, *China Mountain Zhang* (New York: Tom Doherty Associates, 1992), p. 213.

or had them implanted? The real moral of the portability intuition is that for coupled systems to be relevant to the core of cognition, *reliable* coupling is required. As a matter of contingent fact, most reliable coupling takes place within the brain, but there can easily be reliable coupling with the environment as well. If the resources of my calculator or my Filofax are always there when I need them, then they are coupled with me as reliably as we need. In effect, they are part of the basic package of cognitive resources that I bring to bear on the everyday world. These systems cannot be impugned simply on the basis of the danger of discrete damage, loss, or malfunction, or because of any occasional decoupling: the biological brain is likewise at risk of losing problem-solving capacities through lesion or trauma, and occasionally loses them temporarily, in episodes of sleep, intoxication, and emotion, for example. As long as the relevant capacities are generally there when they are required, this seems to be coupling enough.

Moreover, it is not implausible to suppose that the biological brain has in fact evolved and matured in ways which directly factor in the reliable presence of a manipulable external environment. It certainly seems that evolution has favored on-board capacities which are especially geared to parasitizing the local environment so as to reduce memory load, and even to transform the nature of the computational problems themselves. Our visual systems have evolved to rely on their environment in various ways: they exploit contingent facts about the structure of natural scenes,[14] for example, and they take advantage of the computational shortcuts afforded by bodily motion and locomotion.[15] Perhaps there are other cases where evolution has found it advantageous to exploit the possibility of the environment being in the cognitive loop. If so, then external coupling is part of the truly basic package of cognitive resources that we bring to bear on the world.

Language may provide an example. Certainly language seems to be one of the central means by which cognitive processes are extended into the world. Think of a group of people

[14] See e.g. S. Ullman and W. Richards, *Image Understanding* (Norwood, NJ: Ablex, 1984).

[15] See the essays in A. Blake and A. Yuille, eds., *Active Vision* (Cambridge, MA: MIT Press, 1992), and P.S. Churchland, V.S. Ramachandran, and T.J. Sejnowski "A Critique of Pure Vision", in C. Koch and J. Davis, eds., *Large-Scale Neuronal Theories of the Brain* (Cambridge, MA: MIT Press, 1994), pp. 23-60.

sitting around a table brainstorming, or a philosopher who finds that she thinks best by writing, developing her ideas as she goes. In these cases we once again find actively coupled systems accomplishing tasks that might have been achieved, with much more difficulty, within the head of a single agent; and it is language that makes this coupling possible. It may well be that language evolved, at least in part, to provide an extension of our cognitive resources in just this sort of way.

Within the lifetime of an organism, too, individual learning may have molded the brain in ways that actively anticipate the continued presence of those ubiquitous and reliable cognitive extensions that surrounded us as we learned to perform various tasks. Language is again a central example here, as are the various physical and computational artifacts that are routinely used as cognitive extensions by children in schools and by trainees in numerous professions. In these cases the brain develops in a way that complements these external structures, and in so doing learns to play its role with the confines of a unified, densely coupled system. Once we recognize that the environment plays such a crucial role in constraining the way that a cognitive system evolves and develops, we see that extended cognition is no simple add-on extra, but a core cognitive process.

An analogy may be helpful. The extraordinary efficiency of the fish as a swimming device is partly due, it now seems, to an evolved capacity to couple its swimming behaviors to the pools of external kinetic energy found as swirls, eddies and vortices in its watery environment. These vortices include both naturally occurring ones (e.g., where water hits a rock) and self-induced ones (created by well-timed tail flaps). The fish swims by building these externally occurring processes into the very heart of its locomotion routines. The fish and surrounding vortices together constit<ute a unified and remarkably efficient swimming machine.[16]

Now consider a reliable feature of the human environment, such as the sea of words. This linguistic surround envelopes us from birth. Under such conditions, the plastic human brain

[16]See M.Triantafyllou and G. Triantafyllou, “An Efficient Swimming Machine”, *Scientific American* 272, 3, pp. 64-70 (March 1995).

will surely come to treat such structures as a reliable resource to be factored into the shaping of on-board cognitive routines. Where the fish flaps its tail to set up the eddies and vortices it subsequently exploits, we intervene in multiple linguistic media, creating local structures and disturbances whose reliable presence drives our ongoing internal processes. Words and external symbols are thus paramount among the cognitive vortices which help constitute human thought.

4 From Cognition to Mind

So far we have spoken largely about "cognitive processing", and argued for its extension into the environment. Some might think that the conclusion has been bought too cheaply. Perhaps some *processing* takes place in the environment, but the really crucial questions concern *mind*. Everything we have said so far is compatible with the view that truly *mental* states – experiences, beliefs, desires, emotions, and so on—are all determined by states of the brain. Perhaps what is truly mental is internal, after all?

We propose to take things a step further. While some mental states, such as experiences, may be determined internally, there are other cases in which external factors make a significant contribution. In particular, we will argue that *beliefs* can be constituted partly by features of the environment, when those features play the right sort of role in driving cognitive processes. In this way, we can see that the mind truly extends into the world.

First, consider a normal case of belief embedded in memory. Inga hears from a friend that there is an exhibition at the Museum of Modern Art, and decides to go see it. She thinks for a moment and recalls that the museum is on 53rd Street, so she walks to 53rd Street and goes into the museum. It seems clear that Inga believes that the museum is on 53rd Street, and that she believed this even before she consulted her memory. It was not previously an *occurrent* belief, but then neither are most of our beliefs. Rather, the belief was sitting somewhere in memory, waiting to be accessed.

Now consider Otto. Otto suffers from Alzheimer's disease, and like many Alzheimer's patients, he relies on information in the environment to help structure his life. In particular, Otto carries a notebook around with him everywhere he goes. When he learns new information, he writes it down in his notebook. When he needs some old information, he looks it up. For Otto, his notebook plays the role usually played by a biological memory. Today, Otto hears about the exhibition at the Museum of Modern Art, and decides to go see it. He consults the notebook, which says that the museum is on 53rd Street, so he walks to 53rd Street and goes into the museum.

Clearly, Otto walked to 53rd Street because he wanted to go to the museum and he believed the museum was on 53rd Street. And just as Inga had her belief even before she consulted her memory, it seems reasonable to say that Otto believed the museum was on 53rd Street even before consulting his notebook. For in relevant respects the cases are entirely analogous: the notebook plays for Otto the same role that memory plays for Inga. The information in the notebook functions just like the information that‘ constitutes an ordinary non-occurrent belief; it just happens that this information inheres in a physical state located beyond the skin.

The alternative would be to say that Otto has no belief about the matter until he consults his notebook; at best, he believes that the museum is located at the address in the notebook. But if we follow Otto around for a while, we will see just how unnatural this way of speaking is. Otto is constantly using his notebook as a matter of course. It is central to his actions in all sorts of contexts, in just the same way that an ordinary memory is central in an ordinary life. The same information might come up again and again, perhaps being slightly modified on occasion, before retreating into the recesses of his artificial memory. To say that the beliefs disappear when the notebook is filed away seems to miss the big picture in just the same way as saying that Inga's beliefs disappear as soon as she is longer conscious of them. In both cases the information is reliably there when needed, available to consciousness and available to guide action, in just the way that we expect a belief to be.

Certainly, insofar as beliefs and desires are characterized by their explanatory roles, Otto's and Inga's cases seem to be on a par: the essential causal dynamics of the two cases mirror each other precisely. We are happy to explain Inga's action in terms of her occurrent desire to go to the museum and her standing belief that the museum is on 53rd street, and we should be happy to explain Otto's action in the same way. The alternative is to explain Otto's action in terms of his occurrent desire to go to the museum, his standing belief that the Museum is on the location written in the notebook, and the accessible fact that the notebook says the Museum is on 53rd Street; but this would simply complicate the explanation in a superfluous way. If we must resort to explaining Otto's action this way, then we will also have to do so for the countless other actions in which his notebook is involved; in each of the explanations, there will be an extra term involving the notebook. We submit that to explain things this way is to take *one step too many*. It is pointlessly complex, in the same way that it would be pointlessly complex to explain Inga's actions in terms of beliefs about her memory. The notebook is a constant for Otto, in the same way that memory is a constant for Inga; to point to it in every belief/desire explanation would be redundant. In an explanation, simplicity is power.

If this analysis is right, we can even construct the case of Twin Otto, who is just like Otto except that a while ago he mistakenly wrote in his notebook that the Museum of Modern Art was on 51st Street. Today, Twin Otto is a physical duplicate of Otto from the skin in, but the information in the notebook differs. Consequently, Twin Otto is best characterized as believing that the museum is on 51st Street, where Otto believes it is on 53rd. In these cases, a belief is simply not in the head!

This conclusion mirrors that drawn by Putnam and Burge, but we can once again note important differences. In the Putnam and Burge cases, the external features that constitute the differences in belief are distal and historical, so that twins in these cases produce physically indistinguishable behavior. In the cases we are describing, the relevant external features play an active role in the here-and-now, and have a direct impact on behavior. Where Otto walks to 53rd Street, Twin Otto walks to 51st. So there is no question of explanatory irrelevance for this sort of external belief content—indeed, it is introduced precisely because of the central explanatory role that it plays. Like the Putnam and Burge cases, these cases involve differences in reference and truth-conditions, but that is not all they involve: these cases involve differences in the dynamics of *cognition*.[17]

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.

Some will resist this conclusion. An opponent might put her foot down and insist that as she uses the term "belief", or perhaps even according to standard usage, Otto simply does not qualify as believing that the museum is on 53rd Street. We do not intend to debate what is and is not standard usage, however; our broader point is that the notion of belief *ought* to be used in such a way that Otto qualifies as having the belief in question. In all *important* respects, Otto's case is similar to a standard case of (non-occurrent) belief. The differences between Otto's case and Inga's are striking, but they are superficial. By using the "belief" notion in a wider way, it picks out something more akin to a natural kind. The notion becomes deeper and more unified, and is much more useful in explanation.

To provide *substantial* resistance, then, an opponent will have to show that Otto's and Inga's cases differ in some important and relevant respect. But in what deep respect are the cases different? To make the case *solely* on the grounds that information is in the head in one case but not in the other would be to beg the question. If this difference is relevant to a difference in belief, it is surely not *primitively* relevant. To justify the different treatment, we must find some more basic underlying difference between the two.

It might be suggested that the cases are relevantly different in that Inga has more *reliable* access to the information: after all, someone might take away Otto's notebook at any time, but Inga's memory is safer. As before, there is some plausibility in the thought that constancy of

[17] In the terminology of D.J. Chalmers, "The Components of Content", *Mind* (forthcoming), the crucial difference is that where the twins in the Putnam and Burge cases differ only in their *relational* content, Otto and Twin Otto can be seen to differ in their *notional* content, which (Chalmers argues) is the sort of content that governs cognition. Notional content is generally internal to a cognitive system, but in this case the cognitive system is itself effectively extended to include the notebook.

some kind is relevant here: the fact that Otto always uses his notebook played some role in our justifying its cognitive status. If it were instead a book that Otto were consulting as a one-off, we would be much less likely to ascribe him a standing belief. But in the original case, Otto's access to the notebook is very reliable—not perfectly reliable, to be sure, but then neither is Inga's access to her memory. A surgeon might tamper with her brain, or more mundanely, she might have too much to drink. The mere possibility of such tampering is not enough to deny her the belief.

One might worry that Otto's access to his notebook *in fact* comes and goes: he showers without the notebook, for example, and he cannot read it when it is dark. Surely we do not want to ascribe him a belief that comes and goes so easily? Of course we could get around these difficulties by redescribing the situation, but in any case this sort of occasional temporary disconnection does not seem to threaten our claim. After all, when Inga is asleep, or when she is intoxicated, we do not say that her belief disappears. What really counts is that the information is easily available when the subject needs it, and this constraint is satisfied equally in the two cases. If Otto's notebook were often unavailable to him at times when the information in it would be useful, there might be a problem, as the information would not be able to play the role in guiding action that is central to belief; but if it is easily available in most relevant situations, the belief is not endangered.

Perhaps a difference is that Inga has *better* access to the information than Otto does? For example, perhaps Inga's "central" processes and her memory have a high-bandwidth link between them, as opposed to the low-grade connection between Otto and his notebook.[18] But this alone does not seem to be enough to make a difference between believing and not believing. Consider Inga's museum-going friend Lucy, whose biological memory has only a low-bandwidth coupling to her "central" systems, due to nonstandard biology or past misadventures. Processing in Lucy's case might be less efficient, but as long as the relevant information is accessible, Lucy clearly believes that the museum is on 53rd Street. If the connection was too indirect—if Lucy had to struggle hard to retrieve the information with mixed results, or if it had to be retrieved with the aid of a psychotherapist—we might become more reluctant to ascribe the belief, but in these cases we have moved well beyond Otto's situation, in which the relevant information is easily accessible.

Another suggestion could be that Otto has access to the relevant information only by *perception*, whereas Inga has more direct access—by introspection, perhaps. In some ways, however, to put things this way is to beg the question. After all, we are in effect advocating a point of view on which Otto's internal processes and his notebook constitute a single cognitive system. From the standpoint of this system, the flow of information between notebook and brain is not perceptual at all; it does not involve the impact of something outside the system. It is more akin to introspection, or to other sorts of information flow within the brain. The only deep way in which the access is perceptual is that in Otto's case, there is a distinctly perceptual phenomenology associated with the retrieval of the information, whereas in Inga's case there is not. But why should the nature of an associated phenomenology make a difference to the status of a belief? Inga's memory may have some associated phenomenology, but it is still a belief. The phenomenology is not visual, to be sure. But for visual phenomenology consider the Terminator, from the Arnold Schwarzenegger movie of the same name. When he recalls some information from memory, it is "displayed" before him in his visual field (presumably he is conscious of it, as there are frequent shots depicting his point of view). The fact that standing memories are recalled in this unusual way surely makes little difference to their status as standing beliefs.

What the various small differences between Otto's and Inga's cases have in common is that they are all *shallow* differences. In the really deep, essential respects, Otto's case is just like Inga's. The information is reliably there, easily and automatically accessible, and it plays a central role in guiding Otto's thought and action. To focus on the superficial differences would

[18] This might be suggested by some remarks in J. Haugeland, "Mind Embodied and Embedded", Yu-Houng Houng, ed., *Mind and Cognition: Proceedings of the First International Conference on Mind and Cognition* (Taipei: Academia Sinica, to appear).

be to miss the deep way in which for Otto, notebook entries play just the sort of role that beliefs play in guiding most people's lives.

Perhaps the intuition that Otto's is not a true belief comes from a residual feeling that the only true beliefs are occurrent beliefs. Of course, if we take this feeling seriously, Inga's belief will also be ruled out, as will many of the beliefs that we attribute to individuals in everyday life. This would be an extreme view, but it may be the most consistent way to deny Otto's belief. Upon even a slightly less extreme view—the view that a belief must be *available* for consciousness, for example—Otto's notebook entry seems to qualify just as well as Inga's memory. Once dispositional beliefs are let in the door, it is difficult to resist the conclusion that Otto's notebook has all the relevant dispositions.

5 Beyond the Outer Limits

Once we accept the possibility of externally-constituted belief, questions immediately arise about how far we should go. A case like Otto's is perhaps the central case, but all sorts of further "puzzle" cases spring to mind. What of the amnesic villagers in *100 Years of Solitude*, who forget the names for everything and so hang labels everywhere? Does the information in my Filofax count as part of my memory? If Otto's notebook has been tampered with, does he believe the newly-installed information? Do I believe the contents of the page in front of me before I read it? Is my cognitive state somehow spread across the Internet?

We do not think that there are categorical answers to all of these questions, and we do not propose to give them. But to help us understand what is involved in ascriptions of extended belief, we can at least examine the features of our central case that make the notion so clearly applicable there. First, the notebook is a constant in Otto's life—in cases where the information in the notebook would be relevant, he will rarely take action without consulting it.[19] Second, the information in the notebook is directly available without difficulty. Third, upon retrieving information from the notebook he automatically endorses it. Fourth, the information in the notebook has been consciously endorsed at some point in the past, and indeed is there as a consequence of this endorsement. The status of the fourth feature as a criterion for belief is arguable (perhaps one can acquire beliefs through subliminal perception, or through memory tampering?), but the first three features certainly play a crucial role.

Insofar as increasingly exotic puzzle cases lack these features, the applicability of the notion of "belief" gradually falls of. If I rarely take relevant action without consulting my Filofax, for example, its status within my cognitive system will resemble that of the notebook in Otto's. But if I often act without consultation—for example, if I sometimes answer relevant questions with "I don't know"—then information in it counts less clearly as part of my belief system. The Internet is likely to fail on multiple counts, unless I am unusually computer-reliant, facile with the technology, and trusting, but information in certain files on my system may qualify. In intermediate cases, the question of whether a belief is present may be indeterminate, or the answer may depend on the varying standards that are at play in various contexts in which the question might be asked. But any indeterminacy here does not mean that in the central cases, the answer is not clear.

And what about socially-extended cognition? Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle. In an unusually interdependent couple, it is entirely possible that one partner's beliefs will play the same sort of role for the other as the notebook plays for Otto.[20] What is central is a high degree of trust, reliance, and accessibility. In other social relationships these criteria may not be so clearly fulfilled, but they might nevertheless be fulfilled in specific domains. For example, the waiter at my favorite restaurant might act as a repository of my beliefs about my favorite meals (this might even be construed as a case of extended desire). In other cases, one's beliefs might naturally be seen to be embodied in one's secretary, one's accountant, or one's collaborator in a

[19] A "constancy" criterion may suggest that history is relevant. On the other hand, the criterion might be cashed out in purely dispositional terms, as the formulation above suggests. In any case, we are not arguing that no historical feature can be relevant to some state's counting as a belief, as long as the main burden is carried by features of the present.

similar way.[21]

In each of these cases, the major burden of the coupling between agents is carried by language. Without language, we might be much more akin to discrete Cartesian "inner" minds, in which high-level cognition, at least, relies largely on internal resources. But the advent of language has allowed us to spread this burden into the world. Language, thus construed, is not a mirror of our inner states but a complement to them. It serves as a tool whose role is to extend cognition in ways that on-board devices cannot. Indeed, it is not implausible that the explosion of intellectual development in recent evolutionary time is due as much to this linguistically-enabled extension of cognition as to any independent development in our inner cognitive resources.

What, finally, of that most problematic entity, the self? Does the spread of cognitive processes out into the world imply some correlative leakage of the self into local surroundings? It seems so. Most of us already accept that the boundaries of the self outstrip the boundaries of consciousness: My dispositional beliefs, for example, constitute in some deep sense part of who I am. If so, then our previous discussion implies that these boundaries may also fall beyond the skin. The information in Otto's notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that Otto *himself* is best regarded as an extended system, a coupling of biological organism and external resources. To consistently resist this conclusion, we would have to shrink the self into a mere bundle of occurrent states, severely threatening the deep psychological continuity of the self. Far better to take the broader view, and see agents

[20]From the *New York Times*, March 30, 1995, p.B7, in an article on former UCLA basketball coach John Wooden: "Wooden and his wife attended 36 straight Final Fours, and she invariably served as his memory bank. Nell Wooden rarely forgot a name—her husband rarely remembered one—and in the standing-room-only Final Four lobbies, she would recognize people for him."

[21]Might this sort of reasoning also allow something like Burge's extended "arthritis" beliefs? After all, I might always defer to my doctor in taking relevant actions concerning my disease. Perhaps so, but there are some clear differences. For example, any extended beliefs would be grounded in an existing active relationship with the doctor, rather than in a historical relationship to a language community. And on the current analysis, my deference to the doctor would tend to yield something like a true belief that I have some other disease in my thigh, rather than the false belief that I have arthritis there. On the other hand, if I used medical experts solely as terminological consultants, the results of Burge's analysis might be mirrored.

themselves as spread-out into the world.

As with any reconception of ourselves, this view will have significant consequences. There are obvious consequences for philosophical views of the mind and for the methodology of research in cognitive science, but there will also be effects in the moral and social domains. It may be, for example, that in some cases interfering with someone's environment will have the same moral significance as interfering with their person. And if the view is taken seriously, certain forms of social activity might be reconceived as less akin to communication and action, and as more akin to thought. In any case, once the hegemony of skin and skull is usurped, we may be able to see ourselves more truly as creatures of the world.