

(needs footnote ¹⁻⁵ - do you have these?)

(1)

The Extended Mind

Andy Clark & David Chalmers

Department of Philosophy

Washington University

St. Louis, Missouri 63130

dave@twinearth.wustl.edu

March 23, 1995

0. ~~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 reference of our concepts and the truth-conditions of our thoughts "just ain't in the head", 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 ~~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 3D 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} an 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) an identical computational/informational 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,¹ the use of physical re-arrangements of ^{letter} tiles to prompt word recall in Scrabble,² the use of instruments such as the nautical slide rule,³

- 1. McClelland, Rumelhart, Smolensky & Hinton (1986), Clark (1989) ch. 6
- 2. Kirsh (to appear)
- 3. Hutchins (1995)

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 figures (zoids) descend from the top of the screen and must be built into compact walls as the game progresses. A falling zoid must be directed, rather rapidly, to an appropriate slot in the emerging structure. The physical rotation operation is available to players. Kirsh and Maglio (1994, p. 530) calculate that the physical rotation of a zoid 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 zoid ready to fit a slot, but often to help *identify* the shape of a candidate zoid. 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., 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!

2. 3 Active Externalism

The central feature of these cases is that the human organism is ^{linked} ~~coupled~~ 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 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

rice!

add ref.

5

* See e.g. E. Kandel, J. Schwartz & T. Jessell (eds) Principles of Neural Science, 2nd Ed, New York: Elsevier

external features here are just as causally relevant as typical internal features of the brain.⁴

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 why bother? If an isomorphic process were going on in the head, it would seem pointless to characterize it in this cumbersome way. ^{*}In a very real sense, the re-arrangement of tiles on the tray is not part of action; it is part of *thought*.

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 current scientific usage, is tied up with conscious processing. It is widely accepted that all sorts of processes that ^{lie}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. 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 ^{ve}processes in the head is the requirement that cognitive processes be *portable*. Here, we are moved by a vision of what 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

⁴It 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?

add as not a ref?

* Interestingly, Herbert Simon ⁵ once suggested that we view internal memory as, in effect, an external resource upon which ^{the} inner processes operate. "Search in memory" he comments "is not very different from search of the external environment". Simon's position

(Footnote - contd)

representers, as believe, another consists in to address our doubts. For what our arguments

minimally ~~truly~~ demand is parity of treatment between internal & external cases. ~~in which whole the parity can be~~

is brought ^{other} by denoting some ~~case~~ internal & ~~cases~~ ^{by} promoting some external ones.

See H.A. Simon, The Sciences of the Artificial, (Second Edition, MIT Press, Camb. Ma 1981).

Also, see ~~the~~ exchange between M. Stepiak & Simon in W. Clancey, S. Smoliar & M. Stepiak (Eds) Contemplating Minds (MIT Press, Camb. Ma 1994) p. 239-242

de
uncoupled: if the mind were like *that*, it would be constantly gaining and shedding little bits of itself.

We are quite sympathetic to this objection. It seems clear that the brain (or perhaps, in this view, the brain and the body) ^{constitutes} is 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 inconstancy 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 ^{reliably &} available ^{Such cases suggest that even} 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 a central part of thought.}

^{ready to} ^{freely} ~~Even if we take the portability intuition seriously, active externalism need not be rejected.~~ ^{Like wise,} ~~Indeed, to put things this way may be already enough to concede the point in principle.~~ ^{Concede the previous example is} ^{general} We have ^{one} already seen that counting on our fingers has been let in the door, and it ^{becomes} is easy to push things ^{her} further. Think of the old image of the engineer with a slide rule hanging from his belt wherever ^{she} he goes. What if ^{someone} people always carried a pocket calculator, or had them ^{one} implanted? What if ^{see} ~~(((one more example, but not body docks.)))~~ ^{attached}

^{Such} ~~What the portability intuition really tells us is that for coupled systems to be relevant to~~ ^a ^{external} ^{a real part of} the core of cognition, *reliable* coupling is required. As a matter of contingent fact ^{that} 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

What if someone never travelled outside
 their home without an air pollution detector -
 one which issued a warning by emitting a
 high pitched tone? In such cases, the agent will
 be aware only of the results of certain computations,
 and of the specific processes involved. But how is
 this different from a person who
 always knows what day is now, but has no
 further awareness of a ~~computation~~ nature of a
 process which led to her becoming the person? *

All the portability intuition really tells
 us, we assert, is that --

Footnote?

* Compare the following passage from a
 recent science fiction novel:

~~Footnote?~~

* "I am taken to the systems department where
 I am attuned to the system. All I do is jack
 in & the technician instructs the system to attune &
 it does. I jack out & query the time. 10:52. The

Information put up. Always before I could
 only access information when I was jacked
 in, it gave me a sense that I knew what I
 thought + what a system told me, but now,
 how do I know what is system and what is
 Zhang? " M. Mettigh, China Mountain Zhang
 (Tom Doherty Associates; New York, 1992) p. ~~213~~²¹³

We can return to this issue, a
 bondg sense say + could in a final section.

damage, loss

cannot be impugned simply on the basis of the danger of discrete [^]danger or malfunction, or ^{on the basis of any} of the 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.

Indeed, it may even be the case that the biological brain has evolved in ways which factor in the reliable presence of a manipulable external environment. It seems rather plausible to suppose that evolution will favor 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} (animate vision...). SPECIFIC EXAMPLE?—
(attached)

3. 4 From Cognition to Mind

for its extension into

So far we have spoken largely about “cognitive processing”, and argued that ~~cognition~~ extends 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 pursue to

Here, ~~we~~ 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.

Enya Inga

First, consider a normal case of belief embedded in memory. ^{Enya Inga} Karen 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

problems themselves. Our vision systems have
 clearly evolved in ways which are geared to
 the massive & repeated exploitation of
 entirely contingent facts about a structure of
 natural scenes, ^{#1} the ~~availability~~ ^{ready} ~~the~~
 the ability of the ~~visual~~ ^{movable} fovea to
 visit & revisit discrete ~~areas~~ ^{the}
 parts of ~~the~~ the visual array to retrieve information of
 an ~~ade~~ ^{equivalent} (rather build a detailed inner world
 model), ^{#2} the ~~ad~~ ^{consumption} ~~the~~ ^{sketches} ~~of data~~ ~~acquisition~~
 afforded ~~by~~ ^{by} our ~~eyes~~ ^{eyes} ~~to~~ ^{have}
 gross characteristics, ^{#3} ~~judging~~ ^{vision} ~~motion~~ ^{vision}
 & ~~skat~~, is a computational system designed for a
~~own~~ ^{own} & spread the computational load across
 brain, body & world.

Notes

1. See e.g. Ullman, S & Richards, W (1984)
Image Understanding (Ablex, Norwood, N.J)
2. See Ballard, D.H "Animate vision"
Artificial Intelligence 48: 1991, p. 57-86
3. See essays in Blake, A & Yuille, A (Eds)
Active vision (MIT Press, Cambridge 1992) &
 P. Chudhary, V. Ramachandran & T. Sejnowski
 "A critique of pure vision" in C. Koch & J. Davis (Eds)
 Large Scale Neuronal THEORIES OF THE BRAIN (MIT Press, 1994) 23-60

Inga
Karen

the museum. It seems clear that Karen believes that the museum is on 53rd Street, and that she believed this even before she consulted her memory. It was not an ^{previously} *occurrent* belief, but then neither are most of our beliefs. Rather, the belief was sitting somewhere in memory, waiting to be accessed.

Otto. Otto

Now consider ^{Otto} Joe. Joe 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} Joe 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} Joe, his notebook plays the role of a ^{biological} memory. Today, ^{Otto} Joe 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} Joe 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} Karen had her belief even before she consulted her memory, it ^{seems} is reasonable to say that ^{Otto} Joe 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} Joe the same role that memory plays for ^{Inga} Karen. 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} that extends beyond the skin.

The alternative would be to say that ^{Otto} Joe 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} Joe around for a while, we will see just how unnatural this way of speaking is. ^{Otto} Joe 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} Karen's beliefs disappear as soon as she is longer conscious of them. In both cases the

reliably when needed and is to guide

information is always there, available to consciousness and available for action, in just the way that we expect a belief to be.

Certainly, insofar as beliefs and desires are characterized by their explanatory roles, Joe's and Karen's cases seem to be on a par: the essential causal dynamics of the two cases seem mirror each other precisely. We are happy to explain Karen'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 Joe's action in the same way. The alternative is to explain Joe'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 Joe'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 Karen's actions in terms of beliefs about her memory. The notebook is a constant for Joe, in the same way that memory is a constant for Karen; 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 Joe, who is just like Joe except that a while ago he mistakenly wrote in his notebook that the Museum of Modern Art was on 51st Street. Today, Twin Joe is a physical duplicate of Joe from the skin in, but the information in the notebook differs. Consequently, Twin Joe is best characterized as believing that the museum is on 51st Street, where Joe 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

Otto

active role in the here-and-now, and have a direct impact on behavior. Where Joe walks to 53rd Street, Twin ^{Otto}Joe 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*.⁵

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, ^{action guiding} 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}Joe simply does not qualify as believing that the museum is on 53rd Street. We do not propose 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}Joe qualifies as having the belief in question. In all *important* respects, ^{Otto's}Joe's case is similar to a standard case of (non-occurrent) belief. The differences between ^{Otto's}Joe's case and ^{Inga's}Karen'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}Joe's and ^{Inga's}Karen'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.

⁵In the terminology of Chalmers 1995, the crucial difference is that where the twins in the Putnam and Burge cases differ only in their *relational* content, Joe and Twin Joe 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.

It might be suggested that the cases are relevantly different in that ^{Inga} Karen has more *reliable* access to the information: after all, someone might take away ^{Otto's} Joe's notebook at any time, but ^{Inga's} Karen's memory is safer. As before, there is some plausibility in the thought that constancy of some kind is relevant here: the fact that ^{Otto} Joe always uses his notebook played some role in our justifying its cognitive status. If it were instead a book that ^{Otto} Joe 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} Joe's access to the ^{note} book is very reliable—not perfectly reliable, to be sure, but then neither is ^{Inga's} Karen'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} Joe'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 the belief. After all, when ^{Otto} Karen is asleep, or when she is ^{intoxicated} drunk, 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} John'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} Karen has *better* access to the information than ^{Otto} Joe does? For example, perhaps ^{Inga's} Karen's "central" processes and her memory have a high-bandwidth link between them, as opposed to the low-grade connection between ^{Otto} John and his notebook.⁶ But this alone does not seem to be enough to make a difference between believing and ^{not} disbelieving. Consider ^{Inga's} Karen's museum-going friend ^{Nico} Lucy, whose biological memory has only a low-bandwidth coupling to her "central" systems, due to nonstandard biology or past misadventures. Processing

Nico's

Nico

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 Joe's situation, in which the relevant information is easily accessible.

Otto

Another suggestion could be that Joe has access to the relevant information only by perception, whereas Karen 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 Joe'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 Joe's case, there is a distinctly perceptual phenomenology associated with the retrieval of the information, whereas in Karen's case there is not. But why should the nature of an associated phenomenology make a difference to the status of a belief? Karen'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, as there are frequent shots depicting his point of view). The fact that standing memories are recalled in this unusual way makes little difference to their status as standing beliefs.

Otto's Inga's

What the various small differences between Joe's and Karen's cases have in common is that they are all shallow differences. In the really deep, essential respects, Joe's case is just like

Otto's

⁶This might be suggested by some remarks of Haugeland (1993).

Inga's

reliably

Karen's. The information is always there, easily and automatically accessible, and it plays a central role in guiding ^{Otto's} Joe's thought and action. To focus on the superficial differences would be to miss the deep way in which for ^{Otto} Joe, notebook entries play just the sort of role that beliefs play in guiding most people's lives.

Perhaps the intuition that ^{Otto's} Joe'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} Karen'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} Joe's belief. Upon even a slightly less extreme view—the view that a belief must be available for consciousness, for example—^{Otto's} Joe's notebook entry seems to qualify as well as ^{Inga's} Karen's memory. Once dispositional beliefs are let in the door, it is difficult to resist the conclusion that ^{Otto's} Joe's notebook has all the relevant dispositions.

Once we accept the possibility of externally-constituted belief, questions immediately arise about how far we should go. A case like ^{Otto's} Joe'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? What if ^{Otto's} Joe's notebook has been tampered with? Do I believe the contents of the page in front of me before I read it? Can states of my spouse play a role in constituting beliefs of mine? 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. As we move away from the central cases, the notion of belief gradually falls off in its applicability. But we can note some features of ^{ow} the central case that makes the notion seem so clearly applicable there. First, the notebook is a constant in ^{Otto's} Joe's life—in cases where the information in the notebook would be relevant, he will rarely take action without consulting it. 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 status of the "beliefs" in question becomes more unclear. If I am often disposed to take relevant actions without consulting my ^{Spouse} wife, her beliefs are less clearly my beliefs. But an unusually close couple might meet the criteria, at least where some beliefs are concerned. The Internet may fail on multiple grounds (unless I am unusually reliant, facile, 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 in the central cases, with environment tightly coupled with organism, the mental status of these ^{extended} states seems clear.

~~4.5~~

Section on Methodology?

~~S. added~~

~~5 5~~

4. S. Methodology, & Beyond.

~~S. Why care?~~

~~S. Why care?~~

We believe that there is much more at issue here than a somewhat understated decision between two ways to use the concepts of belief & of cognitive process. First, we have claim to have shown that the simple demand for consistency of usage requires parity of treatment for Otto & Inga. But second, we assert that the most fruitful way to achieve such parity is ^{indeed to} ~~to~~ 'upgrade' our concepts & some external structures, rather than to downgrade our understanding of non-occurrent ~~beliefs~~ stored knowledge (so as to derive it of cognitive status). ~~The latter option~~ ^{For} the former option, but not the latter, has desirable ~~theoretical~~ methodological consequences. ~~By For~~ ~~Suppose~~ and helps provide a conceptual foundation for several bodies of recent Cognitive Scientific research. If, as we suppose, our inner computational resources are indeed

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 (see e.g. the growing body of work on the tricks and strategies of so-called Animate vision systems: systems which act so as to maximize the role of external structure in simplifying internal computation (Ballard (1991), Churchland, Ramachandran and Sejnowski (1994), Clark (to appear))).

All of which leads us, at last, to the methodological heart of the matter. To a first approximation it surely matters very little whether we choose to include external structures and manipulations as proper parts of cognitive processes, or merely as additional props and aids. It is, perhaps, a matter of some philosophical interest if (as I have claimed) there exists no good reason to restrict cognitive processes to the bounds of the body, and some good reasons to allow them to range more freely. But beyond a raised eyebrow or two, who would care? Methodologically, however, there is a real battle to be fought. For suppose ~~that inner structure and processes are indeed~~ exquisitely adapted (by evolution or lifetime learning) so as to distribute cognitive tasks, and to solve problems by closely coupled physical interactions with ^{external} ~~further~~ resources. In that case, it makes no sense to study the brain (or brain/body) with as little regard to real environmental interactions as has been ^{in the not-too-distant past} ~~generally~~ customary (with some notable exceptions -- see e.g. Gibson (1979)) ^{This} ~~Thelen and Smith (1994), and the Animate vision literature mentioned earlier)~~. ~~This~~ does not mean that the brain/body is not itself a legitimate target of interest, as conceded earlier. Rather it means that *if* that is your target, you will only understand its true nature by paying close attention to its properties as embedded in some ecologically realistic setting. Treating

Selzys ad

cognitive processes as the kind of things which regularly criss-cross the body/environment

boundary is, ^{we} believe, a crucial step in promoting such a practice. And it is

a step which helps clarify the theoretical foundations of a wide variety of emerging research in areas as diverse as ^{the theory of} Situated Cognition¹, the dynamic systems oriented work on child development², studies of autonomous agents and real-world ~~robotic~~ robotics³, and research on the ~~past~~ ~~over~~ cognitive properties of collectives of agents⁴.

BB1 9.15

SR 9.40

10.00 MW

9.50 COF

1. See e.g. L. Suchman Plans + Situated Actions (Cambridge Univ. Press, Camb. 1987)
2. See E. Thelen + L. Smith A Dynamic Systems Approach to the Development of Cognition + Action (MIT Press, Camb. MA 1994)
3. E.G. R. Beer, Intelligence as adaptive behavior (Academic Press, N.Y.)
4. E. Huetheis, Cognition in The Child (MIT Press, 1995)

What, finally, of that most problematic entity, the self? Does the ~~putative~~ spread of cognitive processes out into the world imply some correlative (and surely unsettling) leakage of the self into the local surroundings? ^I The answer looks to be (sorry!) "Yes and No". No, 22

~~What of that most problematic entity, the self? Does the spread of cognitive processes into the world imply that I am partly out there in the world? If self is determined only by consciousness, then the self may be internal, but it seems fair to say that the self outstrips consciousness: my dispositional beliefs, for example, constitute in some deep sense part of who I am. One can fairly say, then, that the information in Joe's notebook is a central part of his identity as a cognitive agent.~~

Thus it may well be that part of me is out there. In a clear and concrete sense, I am not just constituted by my brain, or my body.

~~because we already conceded that conscious contents supervene on individual brains. But~~

~~Yes, because such conscious episodes are at best snapshots of the self considered as an~~

~~evolving psychological profile. Thoughts, considered only as snapshots of our mental~~

~~activity, ^{is} ^{we are agree} fully explained, I am willing to say, by ^s the current state of the brain. But the~~

~~flow of thoughts, the temporal evolution of ideas and attitudes, these are determined (and~~

~~explained) by the intimate ^{reliable and} continued interplay between the individual brain, and the body and~~

~~world. In constructing this paper -- which is, I promise, a reflection of my attitudes and~~

~~beliefs! -- I relied on constant interactions with external media (paper, pen, computers,~~

~~books). These media allowed ideas to gel and rearrange and mutate in ways which simply~~

~~would not have occurred sitting in a darkened room. The evolving flow of ^{our} my thoughts and~~

~~beliefs is thus not determined or explained simply by some confrontation between ~~my~~ on-~~

~~board reasoning devices and a world of data and inputs. Instead, the world (at times) is~~

~~itself the locus of ^{reliable criterion} operations and transformations, which, ^{operations & transformations} were they to occur in the head, we~~

~~would have no hesitation in calling cognitive processes and in directly associating with some-~~

~~specific psychological profile, ^{an agent} of ~~the self~~.~~

~~But~~
In-sum, let the idea of self fall where it will. ^{our} My concern has been only to argue that

we have good reasons, both philosophical and methodological, to embrace a more liberal

belly ad
 notion of cognitive processes -- one which explicitly allows the spread of such processes across brain, body, world, and artifact. The time is ripe to usurp the hegemony of skin and skull, and to recognize the complex conspiracy that is adaptive success.

**Philosophy/Neuroscience/Psychology Program
 Department of Philosophy
 Washington University
 St. Louis, MO 63130
 e-mail: andy@twinearth.wustl.edu**

dave --

REFERENCES

Ballard, D. (1991). Animate Vision. *Artificial Intelligence* , 48 , 57-86.

Churchland, P., Ramachandran, V., & Sejnowski, T. (1994). A Critique of Pure Vision. In C. Koch & J. Davis (Ed.), *Large-Scale Neuronal Theories of the Brain* . Cambridge, MA: MIT Press.

Churchland, P., & Sejnowski, T. (1992). *The Computational Brain* . Cambridge, MA: MIT Press.

Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing* . Cambridge: MIT Press.

Clark, A. (to appear). Moving Minds: Situating Content in the Service of Real-Time Success. In J. Tomberlin (Ed.), *Philosophical Perspectives* (Vol. 10,).

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception* . Boston: Houghton-Mifflin.

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

Kirsh, D. & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science*, 18, 513-549.

Malone, T., Fikes, R., Grant, K., & Howard, M. (1988). Enterprise: A market-like task scheduler for distributed computing environments. In B. Huberman (Ed.), *The Ecology of Computation* . North-Holland.

McClelland, J., Rumelhart, D., & Hinton, G. (1986). The appeal of Parallel Distributed Processing. In J. McClelland, D. Rumelhart, & P. R. Group (Ed.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (Vol. II, pp. 3-44).

Cambridge: MIT Press/Bradford Books.

Putnam, H. (1975). The meaning of "meaning". In H. Putnam (Ed.), *Mind, Language and Reality* (pp. 215-275). Cambridge: Cambridge University Press.

Rutkowska, J. (1993). *The Computational Infant*. New York: Harvester Wheatsheat.

Searle, J. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

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