Cognitive Science 29 (2005) 135–143
Copyright © 2005 Cognitive Science Society, Inc. All rights reserved.

# Does the World Leak Into the Mind? Active Externalism, "Internalism" and Epistemology

Terry Dartnall

*Computing and Information Technology*
*Griffith University, Australia*

Received 3 February 2003; received in revised form 17 March 2004; accepted 30 April 2004

**Abstract**

One of the arguments for active externalism (also known as the *extended mind thesis*) is that if a process counts as cognitive when it is performed in the head, it should also count as cognitive when it is performed in the world. Consequently, mind extends into the world. I argue for a corollary: We sometimes perform actions in our heads that we usually perform in the world, so that the world leaks into the mind. I call this *internalism.* Internalism has epistemological implications: If a process gives us an empirical discovery when it is performed in the world, it will also give us an empirical discovery when it is performed in the head. I look at a simple example that highlights this implication. I then explore the relation between internalism and active externalism in more detail and conclude by comparing internalism with mental modeling.

*Keywords:* Active externalism; Extended mind; Internalism; Epistemology

## 1. Introduction

Active externalism, also known as the *extended mind thesis,* says that mind extends into the world (Clark, 1997, 2003; Clark & Chalmers, 1998; Dennett, 1996; Donald, 1991; Hutchins, 1995). Clark and Chalmers said that cognitive *processes* extend into the world when we use pen and paper to work something out, or when we use a computer, or even when we use language, which Clark thinks was the first technology. They said that cognitive *states* extend into the world when we use physical objects, or data structures, such as chips or CD-ROMs, as external memory stores that we can consult as needs dictate.

Clark and Chalmers' 1998 article leans heavily on the *parity argument,* which says that if a process counts as cognitive when it is performed in the head, it should also count as cognitive

Requests for reprints should be sent to Terry Dartnall, Computing and Information Technology, Griffith University, Nathan, Queensland 4111, Australia. Preferred e-mail: terrydartnall@hotmail.com; backup e-mail: terryd@cit.gu.edu.au

when it is performed in the world. Since then another argument, the *complementary argument,* has gained the ascendancy. Clark (1998) said, "The argument for the extended mind *turns primarily* [italics added] on the way disparate inner and outer components may co-operate so as to yield integrated larger systems capable of supporting various (often quite advanced) forms of adaptive success" (p. 99). The complementarity argument is elaborated at length in Clark (2003).

In this article I argue for a corollary and complement to active externalism: We sometimes perform actions in our heads that we usually perform in the world, and we typically perform them on inner analogs of external objects. Consequently, the world leaks into the mind. I call this *internalism.* Internalism has epistemological implications: If a process gives us an empirical discovery when it is performed in the world, it will also give us an empirical discovery when it is performed in the head. I begin by outlining the parity argument. Then I look at a simple example of internalism that highlights the epistemological implications. I explore the relation between internalism and active externalism and conclude by comparing internalism with mental modeling.

## 2. The parity argument

Clark and Chalmers (1998) asked us to imagine that we can rotate images of geometrical shapes on a computer screen, either by using a neural implant in our heads or by using a "rotate" button in the world. They say that the implant case is clearly cognitive, so that the button case is as well. Epistemic credit is due where epistemic actions are performed, regardless of whether they are performed in the head or in the world.

This, however, only covers cognitive *processes,* and Clark and Chalmers (1998) admitted that the processes might be in the world while all of our "truly mental states—experiences, beliefs, desires, emotions, and so on" might be in the head (p. 12). To meet this objection they took the parity argument a stage further and argued that cognitive *states* can be constituted partly by features of the environment.

This brings us to the strange case of Otto's notebook. Otto suffers from Alzheimer's disease. He hears that there is an exhibition at the Museum of Modern Art. He consults his notebook, which says that the museum is on 53rd Street. He walks to 53rd Street and goes to the museum. Clark and Chalmers (1998) said that the notebook plays the same role for Otto that biological memory plays for the rest of us. It just happens that "this information lies beyond the skin" (p. 13). Otto believed that the museum was on 53rd Street before he looked it up, courtesy of the functional isomorphism between the notebook entry and a corresponding "entry" in biological memory. When we remember something that was stored in long-term biological memory, we say that we knew it before we consciously recalled it. Otto's notebook plays the same role for him that long-term biological memory plays for the rest of us: Otto knew the address of the Museum of Modern Art before he consulted his notebook.

## 3. Internalism and epistemology

I argue for internalism and its epistemological implications by looking at a simple example. You walk into a room and see a partially completed jigsaw puzzle on the table. You look at the

puzzle and leave the room. You then mentally rotate one of the pieces and discover where it fits into the puzzle. R. M. Shepard and associates showed in a series of classic experiments conducted in the early 1970s that we can perform operations such as these (Cooper & Shepard, 1973; Shepard & Metzler, 1971).[1]

You have now discovered something new—where the piece fits into the puzzle. But how did you discover it? You did not discover it by straightforward empirical discovery, because you did not have access to the puzzle at the time. Nor did you remember it. You could only have remembered what you knew when you were in the room, and you did not know where the piece fits into the puzzle when you were in the room. I believe that you discovered it by performing an operation in your head that you would normally have performed in the world. Consequently, I think that you made an empirical discovery by performing an operation in your head.

Most people find this claim counterintuitive and believe, instead, that you derived the knowledge from what you already knew. There is an obvious reason for this belief. You did not have access to the puzzle after you left the room, so it seems that you must have derived the knowledge from what you already knew when you were in the room.

One version of this theory is that you derived the knowledge deductively and *then* imagined the piece fitting into place. That is, you remembered the state of the puzzle, *inferred* that the piece would fit, and on this basis imagined it fitting into place. The imagery was epiphenomenal: It played no functional or causal role in the discovery.

I think that there are situations in which this kind of claim is plausible. If I know that lead is soft, I can imagine what will happen when I hit a piece of lead with a hammer. But I can derive this conclusion without imagining anything at all. I do not need to imagine the hammer actually flattening the lead. So it might be that in this case I *first* make the inference and *then* imagine the conclusion (the hammer flattening the lead).

Could you have been inferencing like this in the jigsaw puzzle case? I do not think so. I can make the inference about the lead because there is a covering law about lead (it is soft and will flatten when hit with a hard object) and relevant information about what we are going to do to it (hit it with a hard object). An anonymous referee has pointed out that there are two covering laws in the jigsaw puzzle case—a figure at Orientation 1 with shape S entails a figure at Orientation 2 with shape S, and if there is a match between figure and whole, then the figure fits the whole. I do not think that this gets us very far. It only says that a figure has the same shape when its orientation changes. It does not help us to say whether the figure will fit when it is rotated. We want to know what the figure *looks like* when it is rotated, not merely that its shape does not change.

If the rotation was epiphenomenal, then you first inferred the fit and then imagined the rotation. I do not know whether you could have inferred the fit without mentally rotating the piece, but there does seem to be an onus-of-proof situation here. It certainly *seems* that rotating the piece plays a role in discovering the fit, so the onus of proof is on the inferentialist to show that this is not the case.

Another move that we can make is to erode the intuition that underlies the inferentialist's position. This intuition is that you must have derived the knowledge from what you already knew, because you did not have access to the puzzle after you left the room. So imagine two different cases. In the first case you rotate the piece manually. You have now discovered something that you did not know until you performed the rotation—you have made an empirical discovery. It seems to me that there is no epistemological difference between rotating the piece

manually and rotating it mentally. In one case you rotate it manually and find where it fits. In the other you rotate it mentally and find where it fits. Whatever we say about the one we will have to say about the other. Because the manual case gives us an empirical discovery that is not derived from previous knowledge, we should say that the mental case gives us an empirical discovery that is not derived from previous knowledge.

In the second case you rotate the piece in your mind while you are still looking at it (while you are still in the room). This is performing an operation in your mind that you would normally perform in the world. This operation does not explicitate or "tease out" something that you already implicitly know. It enables you to acquire new empirical knowledge. If we say that it explicitates what you already know, we will have to say that a great deal of the knowledge that we apparently acquire through empirical discovery is acquired by explicitating what we already know, so that it is not empirical discovery at all. Now, if rotating the piece in your mind while you are looking at it yields an empirical discovery, what difference is there if you perform the same operation when you are *not* looking at it? This will give you an empirical discovery as well.[2]

Internalism might cast light on a long-standing epistemological issue. Classical rationalism says that we can acquire knowledge about the world through thought and reflection alone, whereas classical empiricism says that we can only acquire it through experience. Internalism says that we can acquire knowledge about the world through thought and reflection, in the sense that we can acquire it through the offline deployment of our sensory abilities. This might be acceptable to classical empiricism, because it talks about acquiring knowledge through experience and the senses, and it might be acceptable to classical rationalism because it says that we can acquire knowledge about the world when our senses are not actively engaged with the world.

## 4. Internalism and active externalism

I now explore the symmetry between internalism and active externalism in more detail. Active externalism says that mind extends into the world through cognitive processes and cognitive states. Internalism most obviously complements this claim in the area of cognitive *processes.* Active externalism says that when we use a rotate button to rotate shapes on a screen, we are performing a cognitive act. Internalism says that when we mentally rotate the jigsaw piece, we are performing an action in our minds that we would normally perform in the world.

The belief that cognitive *states* extend into the world draws inspiration from Brooks' claim that the world "serves as its own best model" (Brooks, 1991, p. 145; see also, Clark, 2003). Clark talked about "relying largely upon the persistent physical surroundings themselves to act as a kind of enduring, external data-store: an external 'memory' available for sampling as needs dictate" (p. 68). He said that we use the world as a memory store even in our most mundane moments. We saccade around a room and foveate onto features of objects, returning to the same features time and again. We do this because *that is where the information is.* We know it is there and return to it time and again—as needs dictate.

I think we can take this a stage further. I think we have inner analogs of objects and states of affairs in the world, which we carry around in our heads and consult as needs dictate. A fairly

weak reason for believing this has to do with symmetry. We perform cognitive actions in the world and perform actions in our heads that we would normally perform in the world. That is the first symmetry. We use the world as an external data store and consult it as needs dictate. If the symmetry carries over we will have inner analogs of external data stores, which we carry around in our heads and consult as needs dictate.

A stronger reason is that there is a problem with using external objects as memory stores. They are not portable. We might use this blackboard or these faces as memory stores when they are in front of us, but they are difficult things to carry around. Having inner analogs of them would overcome this problem and free cognition from the here and now, the context, and the moment.

Robert Goldstone (personal communication, April 29, 2004) pointed out a nice variation on this theme. Shepard (1984) argued that minds have evolved so that their internal constraints match external constraints. This enables them to predict what will happen in the external world. In his discussion of why animals have built-in 24-hr circadian cycles, he said, "Even though it is correlated with the waxing and waning of daylight, this periodicity has become internalized so that it continues autonomously in the absence of the correlated stimulus, freeing the animal from a direct dependence on that stimulus" (p. 422).

Another argument for internal analogs is this: If we perform operations in our minds that we would normally perform in the world—more specifically, if we perform operations in our minds *in the absence of the original object*—then we must have an inner, remembered analog to perform the operation on. In the case of the jigsaw puzzle, we perform an operation in our minds that we would normally perform in the world. We say that we rotate the piece in our minds. But what do we really rotate? The answer seems to be: an inner analog of the jigsaw piece.

Things might not be what they seem, however. Perceptual activity theorists, such as Thomas (1999), hold that, rather than storing inner analogs of the external world, we *generate* them by running our perceptual abilities offline. The general idea is that when we imagine something, such as a cat, we employ the schema that we employ in perceiving a cat, but now we employ it in the absence of the cat. When we imagine the jigsaw piece, we employ the schema that we employ when we actually perceive it. This enables us to generate an image of the piece.

This is a different account to the stored analog account—but it is not that different. Both accounts say that we can consult inner images of the world on a need-to-know basis, to acquire information that was not explicitly stored. The inner analog account says that the image was stored. The perceptual activity account says that it was generated.

## 5. Active externalism and mental models

The relation between internalism and mental modeling depends on what we take internalism to be. I have characterized it as the claim that we perform operations in our heads that we would normally perform in the world and that we typically perform them on inner analogs of external objects. This corresponds to what we might call "strong" mental modeling. I contrast this with weak mental modeling, which is performing operations on things in our heads that we would normally perform *in our heads* on things in the world. The crucial difference is

whether we perform the original action in our heads or in the world. Perceiving something and then imagining it is a case of weak mental modeling, because we perform both operations in our heads. Manipulating something in the world and then manipulating an inner analog is strong mental modeling. I draw this distinction because weak modeling does not import in-the-world operations into our minds and so does not give us interesting and convincing cases of the world leaking into the mind.

An example of weak mental modeling is Kosslyn's (1980, 1994) claim that we construct "quasi-pictures" or "surface representations" in a visual buffer on the basis of information stored as deep representations in long-term memory. Once the quasi-picture has been constructed, it is available as a conscious image that can be read and interpreted in the visual buffer to extract information that was not explicitly represented in long-term memory. If we want to know whether frogs have lips or foxes have pointed ears, we construct an image in the visual buffer and scan it for the required information. In such a way, visual imagery enables us to retrieve information that is not explicitly encoded at the deep level.

There are similarities between Kosslyn's (1980, 1994) pictorialism and internalism. Kosslyn said that when we scan inner images we employ perceptual processing mechanisms that we normally employ in perceiving the world, and we apply them to inner entities that in some ways behave like things in the world. Most obviously, we can scan the images for information, just as we can scan things in the world for information. Kosslyn wrote that the visual buffer is a stage in perceptual information processing that consists of retinotopic maps of the brain's occipital cortex (Kosslyn, 1994). Consequently, he wrote that the contents of the visual buffer when we scan visual images are much the same as its contents when we perceive the external world.

But there are differences as well. Internalism says that we perform actions in our heads that we normally perform in the world. Kosslyn's (1980, 1994) claim is weaker than this: When we scan inner images we employ perceptual processing mechanisms that we normally employ in processing information about the world. This "normal employment," however, takes place in our heads, not in the world. We perform operations in our heads on things in the world (frogs and foxes) and perform the same operations on things in our heads (images of frogs and foxes). This is weaker than internalism, which says that we perform operations in our heads that we normally perform in the world.

Most studies of mental modeling are studies of weak modeling. Consider Gentner and Gentner's study of mental models of electricity (Gentner & Gentner, 1983). According to this, we internally "run" an electrical circuit: We imagine something that we could in principle perceive in the world, such as a "visibly working model" of an electrical circuit. This is weak modeling, because we imagine something that we could see in the world, and both imagining and seeing take place in the head.

Now consider a case of strong mental modeling. Rick Grush (in press) developed a mental modeling framework in terms of emulator theory that aims to synthesis "a great deal of motor control and motor imagery work" as well as aspects of visual imagery and visual perception. What he calls the *emulation theory of representation* says that the brain constructs neural circuits that act as models of the body and environment:

> During overt sensorimotor engagement these models [of the body and environment] are driven by efference copies, in parallel with the body and environment, in order to provide expectations of the

> sensory feedback, and to enhance and process sensory information. These models can also be run off-line to produce imagery, estimate outcomes of different actions, and evaluate and develop motor plans. (Abstract)

Perception involves "a content-rich emulator-provided expectation that is corrected by sensation" (Grush, in press, sect 6.3.1). We can then use efference copies of motor commands to run the emulator offline. This will provide us with similar content, now in the form of mental imagery.

In my commentary on Grush's article (Dartnall, in press) I suggested that the continuity between online and offline emulation explains the jigsaw puzzle case. When we rotate the piece manually, we are running the emulator online to anticipate, fill in, and enhance sensory feedback. When we rotate it mentally (after we have left the room) we are running the emulator offline, using efference copies of motor commands. The emulator gives us similar content to the content it gave us when we were running it online, but now without any input or feedback from the world. Under these circumstances we really are performing operations in our heads that we would normally perform in the world. Grush agreed. He said:

> Dartnall's suggestion is that the world can leak into the mind. I think I agree entirely with this suggestion, and in fact in Grush (2003, Section 6) I discuss this a bit. The basic idea is that emulators in the brain are typically if not always constructed and maintained as a function of observing overt interaction; their ability to represent the target system is in some strong sense dependent on the target system itself, and the details of the organism's (or some other entity's) interaction with it. (Grush, in press)

Identifying internalism with strong mental modeling pulls together two bodies of research: active externalism, which says that we perform cognitive operations in the external world and use the external world as an information store, and mental modeling, which says (depending on the variety) that we run models of the world in our minds or perform operations in our minds that we would normally perform in the world.

## 6. Other issues

Active externalists are at pains to point out that we use objects and states of affairs as external data stores that we can consult as needs dictate. Internalism says that we can have inner analogs of external objects, which we can consult as needs dictate. The economies of storage carry over for internalism enables us to retrieve knowledge that, in a sense, we carry at zero computational cost. I think there is a mystery here. We retrieve knowledge by scanning and manipulating inner analogs, just as we acquire knowledge by scanning and manipulating things in the world. But how can we retrieve knowledge from inner analogs if it is not encoded in the analogs? Perhaps we should say that we "acquire" or "generate" the knowledge, but I do not see how changing the vocabulary can solve the problem.

Nor is it clear why we import the world into our minds. Grush says that for reasons of motor efficiency we need to be able to anticipate and emulate the world, and we can then run this emulation offline. I have suggested that inner analogs free us from the here and now, the context, and the moment. They enable us to carry the world within us so that we can consult it as needs dictate. These are consistent claims. We might have developed the ability to anticipate and em-

ulate the world, and the inner copies that this made possible might have been the beginning of memory.

The notion of an "inner analog" is a nagging concern, because an inner analog is not the same as an inner object, and if we have only inner analogs, then the world might not leak into the mind in the way I have suggested. There are two points here. First, this would still leave the *processes* claim intact, that we perform operations in our minds that we usually perform in the world. Second, it may be significant that we have trouble making sense of objects leaking into the mind, just as we have trouble with the idea of cognitive states leaking into the world. Active externalism says that Otto's state of believing that the museum is on 53rd Street is really *out there in his diary.* We understand what it means to say that we sometimes perform operations in the world that we would usually perform in our minds, and sometimes perform operations in our minds that we would usually perform in the world. It is more difficult to understand how mental states can be in the world and how physical objects can intrude into our minds. It may be that these claims stand or fall, or need to be attenuated, together. If that is the case, the symmetry between internalism and active externalism will remain.

## Notes

1. Finke, Pinker, and Farah (1989) showed that we can rotate objects in our minds and assign a different meaning to them on the basis of the rotation. For example, we can rotate the letter *D* and add it to an upright *J* to get an image of an umbrella.
2. Andy Clark and Dave Chalmers have independently suggested (personal communication, November 21, 2002 and January 15, 2003, respectively) that the argument might go the other way: It might be argued that the mental rotation case involves inference, so that the manual case does as well. I do not see how the manual case can involve inference.

## Acknowledgments

Thanks to Lee Bowie, Dave Chalmers, Andy Clark, John Connolly, Jay Garfield, Howard Skulsky and the PATF group at Smith College for comments on an earlier version of this article. Special thanks to Robert Goldstone for numerous constructive suggestions.

## References

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence, 47,* 139–159.

Clark, A. (1997). *Being there: Putting brain, body and world together again.* Cambridge, MA: MIT Bradford.

Clark, A. (1998). Review symposium of *Being there. Metascience, 7,* 70–104.

Clark, A. (2003). *Natural-born cyborgs.* Oxford, England: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York: Academic.

Dartnall, T. H. (in press). Epistemology, emulators, and extended minds: Peer commentary on Grush (forthcoming). *Behavioral and Brain Sciences.*

Dennett, D. (1996). *Kinds of minds.* New York: Basic Books.

Donald, M. (1991). *Origins of the modern mind.* Cambridge, MA: Harvard University Press.

Finke, R., Pinker, S., & Farah, M. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science, 13,* 51–78.

Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 99–129). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Grush, R. (2003). In defense of some "Cartesian" assumptions concerning the brain and its operation. *Biology and Philosophy, 18,* 53–93.

Grush, R. (in press). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences.*

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

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate.* Cambridge, MA: MIT Press.

Shepard, R. N. (1984). Ecological restraints on internal representation: Resonant kinematics of perceiving, imagining, thinking and dreaming. *Psychological Review, 91,* 417–447.

Shepard, R. N., & Metzler, J. (1971, February 19). Mental rotation of three-dimensional objects. *Science, 171,* 701–703.

Thomas, N. (1999). Are theories of imagery theories of imagination? *Cognitive Science, 23,* 207–245.