

Attacking the Bounds of Cognition

Richard Menary

Recently internalists have mounted a counter-attack on the attempt to redefine the bounds of cognition. The counter-attack is aimed at a radical project which I call “cognitive integration,” which is the view that internal and external vehicles and processes are integrated into a whole. Cognitive integration can be defended against the internalist counter arguments of Adams and Aizawa (A&A) and Rupert. The disagreement between internalists and integrationists is whether the manipulation of external vehicles constitutes a cognitive process. Integrationists think that they do, typically for reasons to do with the close coordination and causal interplay between internal and external processes. The internalist criticisms of the manipulation thesis fail because they misconstrue the nature of manipulation, ignore the hybrid nature of cognition, and take the manipulation thesis to be dependent upon a weak parity principle.

Keywords: Extended Mind; Cognitive Integration; Cognitive Process; Intrinsic Content; Conventional Content; The Parity Principle

1. Introduction

Recently internalists (Adams & Aizawa, 2001, 2006; Rupert, 2004) have mounted a counter-attack on the attempt to redefine the bounds of cognition. Their counter-arguments are aimed at the extended mind hypothesis, which, as Andy Clark has recently put it, is the view that “the material vehicles of cognition can be spread out across brain, body and certain aspects of the physical environment itself” (Clark, 2005, p. 1).

However, I think that the extended mind hypothesis is part of a more radical project which I call “cognitive integration,” which is the view that internal and external vehicles and processes are integrated into a whole. It is this more radical project that Clark and others are really engaged in. Cognitive integration can be

Correspondence to: Richard Menary, Philosophy School of Humanities, University of Hertfordshire, de Havilland Campus, Hatfield, Hertfordshire AL10 9AB, UK. Email: r.menary@herts.ac.uk

defended against the internalist counter arguments of Adams and Aizawa (henceforth A&A) and Rupert, and I shall endeavor to show how.

2. Defining Cognitive Integration

Cognitive integration is committed to the following theses:

- The manipulation thesis.
- The hybrid mind thesis.
- The transformation thesis.
- The cognitive norms thesis.

The manipulation thesis. To explain how this integration works we need to place the cognizer within an environment, where various vehicles are manipulated. On this account we must recognize that the agent often completes a “cognitive task”¹ by manipulating vehicles in the environment—and that these external vehicles are often classical vehicles. The agent sometimes completes cognitive tasks individually by manipulating external classical vehicles—obvious examples being written and spoken language and mathematical symbols—and sometimes by co-operating with other cognitive agents in the shared environment (Clark & Chalmers, 1998; Hurley, 1998; Menary, in press; Rowlands, 1999).

The hybrid mind thesis. Furthermore, cognition is understood as the integration of internal non-classical vehicles and processes—like those commonly found in neural networks—with external classical vehicles and processes. Controversially, on this account, the process of remembering is constituted by the integration of internal and external vehicles and processes which complement one another in the completion of a cognitive task (Clark, 2001, 2003; Menary, in press; Rowlands, 1999; Sutton, 2006).

The transformation thesis. We are able to manipulate external vehicles in the completion of cognitive tasks because we have learnt to do so; this is to say that our abilities to complete cognitive tasks have been transformed by learning the *practice* of manipulating external vehicles. We engage in cognitive practices, manipulating external vehicles, to complete cognitive tasks and this is the mark of the cognitive (Menary, in press; Vygotsky, 1978; Wertsch, 1985).

The cognitive norms thesis. Explaining the integration of the internal and external involves both a dynamical account of the reciprocal causal interaction between internal and external vehicles and processes; and an account of how we learn to manipulate external vehicles by applying the relevant cognitive norms for manipulating external vehicles. For example we learn how to manipulate external mathematical notations to complete a long multiplication, because we learn the norms by which we manipulate mathematical notations. Cognitive norms and cognitive practices are both cognitive and distinguishable from other norms and practices *because* they are aimed at the completion of cognitive tasks (Menary, in press).

The main cognitive integrationist thesis that comes under attack from A&A and Rupert is what Mark Rowlands (1999) calls the “manipulation thesis”:

[C]ognitive processes are not located exclusively in the skin of cognising organisms because such processes are, in part, made up of physical or bodily *manipulation* of structures in the environments of such organisms. (p. 23)

So notice that there is a locational aspect to this claim, that cognitive processes are not exclusively located in the body, underwritten by a constitutive claim that cognitive processes are, in part, constituted by the bodily manipulation of structures in the environment. A&A (2001, 2006) and Rupert (2004) do not deny that manipulating external vehicles, like mathematical symbols and diagrams are important to our ability to complete cognitive tasks; they just deny that the manipulation of these external vehicles constitutes a cognitive process. The real disagreement between internalists and integrationists is whether the manipulation of external vehicles *constitutes* a cognitive process. Integrationists think that they do, typically for reasons to do with the close coordination and causal interplay between internal and external processes. A&A defend a traditional cognitive internalism, by which the bounds of cognition end at the skull. However, I’ll suggest, the internalist criticisms of the manipulation thesis fail because they misconstrue the nature of manipulation, ignore the hybrid nature of cognition and take the manipulation thesis to be dependent upon a weak parity principle.

Before turning to their criticisms I will elaborate briefly on the manipulation thesis. There are at least three classes of manipulation:

1. Biological cases of coupling such as extended phenotypes and animate vision (biological coupling).
2. Using the environment as its own representation, obviating the need for internal representations—as in Tetris (epistemic actions).
3. And most importantly, the manipulation of external representational and notational systems according to certain normative practices—as in mathematics (cognitive practices).

Biological coupling can most easily be understood in terms of extended phenotypes (Dawkins, 1982), where the boundary between phenotype and environment breaks down. Nature is quite happy to incorporate the environment as part of the organismic process. Such processes often involve a reciprocal causal relation where two mechanisms co-ordinate in the completion of some task. Take for example phonotaxis in crickets (Webb, 1994) where the song of the male cricket is iterated in such a way as to allow the primitive neuronal “tracker” mechanisms in female crickets to direct the female cricket towards the source of the male song; and, of course, the coordination of producer and consumer mechanisms in Millikan’s (1993, 2004) justly famous example of bee dances. Theoretical treatments of such coordinated reciprocal causation in humans can be found in ecological approaches to perception (Gibson, 1979), animate vision (Ballard, 1991) and sensorimotor contingencies (O’Regan & Noë, 2001).

The second class of manipulations have been dubbed “epistemic actions” by Kirsh and Maglio (1994). An epistemic action involves manipulating the environment to bring about a better state in a problem-solving/planning task. Thus, an epistemic action is not simply a behavioral output that results from internal planning, it is part of the processing involved in planning/problem-solving.

The classic case of the third class of manipulations is Rumelhart and McClelland’s (1986) example of using pen and paper to complete a mathematical algorithm. Performing long multiplication involves mastery over a notational system, which involves cognitive norms for manipulating those notations when completing cognitive tasks.

The extended mind hypothesis contributes to the radical project of integrating the internal and external, because it provides a way of thinking about the dynamical integration of internal and external vehicles and processes. The internalist criticisms focus only on the external vehicles and processes and ignore their integration with the internal. If integrationists were committed to the claim that external vehicles were made cognitive or mental just because we are causally linked to them, then extended mind style arguments would be very weak. Integrationists do not claim this and the internalist criticisms that follow apply only to a weakened version of cognitive integration.

In the rest of the paper I defend some of the central claims of cognitive integrationists against the internalist arguments of A&A and Rupert.

3. The “Coupling-Constitution Fallacy”

The first attack centers on the argument that cognitive processes could be constituted by causal couplings, or manipulations of external vehicles. Internalists argue that coupling, or manipulation, relations are distinct from constitutive relations. A&A (2006) explain the fallacy as follows:

When Clark makes an object cognitive when it is connected to a cognitive agent, he is committing an instance of a “coupling-constitution fallacy.” This is the most common mistake that extended mind theorists make. The fallacious pattern is to draw attention to cases, real or imagined, in which some object or process is coupled in some fashion to some cognitive agent. From this, one slides to the conclusion that the object or process constitutes part of the agent’s cognitive apparatus or cognitive processing. If you are coupled to your pocket notebook in the sense of always having it readily available, use it a lot, trust it implicitly, and so forth, then Clark infers that the pocket notebook constitutes a part of your memory store. (p. 46)

It follows that the coupling of an object *X* to a cognitive agent does not entail that *X* is part of the agent’s cognitive apparatus. However, the intrinsic nature of *X* does determine whether or not it is cognitive. The intrinsic nature of pens, paper and other artifacts is not cognitive. The only entities that are intrinsically cognitive are brains. Therefore external artifacts are not part of cognition.

So such critics take it that cognitive integrationists think that external vehicles are cognitive *because* they are coupled to individuals. They usually cite Clark and Chalmers (1998; henceforth C&C) as committed to such a view. I think that this is a misinterpretation, rooted in C&C's use of the parity principle to motivate examples of extended cognition.

The parity principle can be simply stated like this: if this external process were located in the skull, we'd call it cognitive. Therefore, even though it isn't in the head we *ought* to call it cognitive.² This is supposed to be an intuition pump, a way of overcoming Cartesian prejudices. It is not, in itself, an argument for the extended mind. However it has become something of an albatross around Clark's neck. Internalists have seized upon the parity principle, attacking a flawed comparative version which, schematically, can be stated as: if an external process/vehicle *X* plays a *similar* role to an internal process/vehicle *Y* (where *Y* is cognitive), then *X* is cognitive. I don't think that C&C's version of the parity principle is helpful as currently framed, nor as caricatured by the internalists.

A&A and Rupert's arguments against the manipulation thesis will be shown to go wrong, because they are aimed at the flawed version of the parity principle. However, the manipulation thesis does not depend upon the parity principle. It follows that cognitive integrationists should drop the flawed version of the parity principle as a motivational tool.

4. Response to the Coupling-Constitution Fallacy

The coupling-constitution fallacy is a misunderstanding of the nature of the coupling/manipulation relation as formulated by cognitive integrationists. A&A (2006) claim that the causal coupling of *X* to *Y* does not make *X* a part of *Y*. The alleged fallacy assumes something like the following picture: an external object/process *X* is causally coupled to a cognitive agent *Y*. The Otto example fits this picture: a notebook coupled to a discrete cognitive agent, whereby the notebook becomes part of the memory system of that agent because it is coupled to the agent. Cognitive integrationists should resist this picture. It is a residual form of internalism, because it assumes a discrete, already formed, cognitive agent. And this is precisely the picture we are arguing against. If we accept the picture of a cognitive agent as implementing a discrete cognitive system, before they ever encounter an external vehicle, then we will have accepted the very picture of cognition we set out to reject. This does not fit with the aim of cognitive integration, which is to show how internal and external vehicles and processes are integrated in the completion of cognitive tasks (such as remembering the location of MOMA).

For the cognitive integrationist the picture is like this: my *manipulation* of the notebook and my brain processes together constitute a process of remembering. In cases like these, the process of remembering cannot be described exclusively in terms of biological memory or solely in terms of the manipulation of external representations, because it is a hybrid process.

Schematically: *X* is the manipulation of the notebook *reciprocally* coupled to *Y*—the brain processes—which together constitute *Z*, the process of remembering. Once we have this picture, it is easy to see that A&A have distorted the aim of cognitive integration. The aim is not to show that artifacts get to be part of cognition just because they are causally coupled to a pre-existing cognitive agent, but to explain why *X* and *Y* are so coordinated that they together function as *Z*, which causes further behavior.³

Take this analogous example: the input layer of units in a feed-forward neural network is coupled to the hidden layer of units, but nobody thinks that this makes the input layer part of the hidden layer. However, the coupling of the input units to the hidden layer units does make them part of a wider system, the neural network.

The manipulation of external vehicles plays a role in completing a cognitive task, as does the processing of internal vehicles—these roles may be different, but they are complementary. The processing of the task is understood in terms of the integration of internal and external processes. Although there will be cases where we will be more like Inga and rely upon biological memories, there will be many cases where we will be more like Otto and recall a memory by manipulating external vehicles (stored in notebooks, or PDAs, etc).

I agree with A&A and Rupert that the hybrid nature of cognition has to be established independently before a project like cognitive integration can be engaged in—such an argument is beyond the scope of this paper. However, I am working under the assumption that examples of hybridity and the complementary dance of internal and external processes have been quite fully described in other sources (Clark, 1997; Hutchins, 1995; Menary, in press; Rowlands, 1999; Wheeler, 2005).

A&A and Rupert have misconstrued the nature of the manipulation thesis, assuming that it is simply a case of causal coupling of a pre-existing cognitive agent to an artifact. It is not. It is also not the case that the parity principle is simply a case of similarity between the internal and external. External vehicles and their manipulation may be really quite different from internal ones, it is the coordination (or reciprocal coupling) of internal and external processes that together produce behavior and there may be equivalent cases where the coordination of processes is solely internal and still leads to the same ends.

5. The Intrinsic Content Condition

A&A (2001, 2006) stipulate that if a process is to be counted as cognitive it *must* involve at least some intrinsic/non-derived content. Any cognitive process *must* involve at least some vehicles with intrinsic/non-derived content (henceforth intrinsic). Note that they do not make the strong claim that cognition *must and can only* be the processing of representations with intrinsic content. A&A robustly state that “Clearly, we mean that if you have a process that involves no intrinsic content, then the condition rules that the process is non-cognitive”

(A&A, 2006, p. 48). If we take it that external vehicles do not have intrinsic content, their content is conventionally determined, then it appears that they cannot be cognitive. There is an important difference between brain states that have their contents intrinsically and words and pictures, which get their contents through social convention. A&A (2001, p. 49) take it that if you are committed to a representationalist and/or language of thought theory of cognition, then you need a theory of intrinsic content. This seems right to me.

A&A (2006) are also very clear that they think that there are mental representations of things like trees, rocks, birds and grass and that there are also mental representations of words, stop signs, warning lights and gas gauges. The mental representations of natural objects are prime cases of intrinsic content. Their contents are fixed by satisfying naturalistic conditions on meaning—as you might find in Fodor’s (1990) asymmetric dependency account, or Millikan’s (1984, 2004) biosemantic account or Dretske’s (1988) indicator semantics. However, words, stop signs, etc. get their contents through social conventions. Does it follow that mental representations of these items have the same conventional contents?

A&A reject this possibility. In replying to A&A, Clark (2006) suggests that a mental representation of a diagram of Euler circles has conventional content. A&A make a distinction between Euler circles that have a conventional meaning and images of Euler circles that have their meanings determined by naturalistic conditions. So, convention determines the meaning of Euler circles and the word ‘dog’.

But that does nothing to show that it is not the satisfaction of some set of naturalistic conditions on non-derived content that get something in the head to have the meanings of “dog,” a stop sign, a white flag, and a warning light. (A&A, 2006, p. 50)

It follows that the meanings of mental representations and conventional signs converge, even though their meanings have been fixed via different channels—natural and conventional. A&A allow Clark the latitude to claim that the content of the image of an Euler circle is dependent upon the social convention, but not that this convention *constitutes* the content of the mental image. They say:

The dependence of meaning of the mental image of intersecting Euler circles on the social contrivance regarding the intersection of Euler circles is just like the dependence of the meaning of a mental representation of a car on the contrivance of a car. Had the car not been invented, there would not have been mental images of cars. (p. 51)

A&A are telling us that the only difference between a mental image of a tree and a mental image of a car is that the former is of a natural kind and that the latter is of an artifact. The difference in object type does not affect the way that the content of images gets determined. Hence, Clark is wrong to assume that some mental images get their contents determined conventionally. The image of an Euler circle is determined by causal correlation, or function of indication, or whatever, but the Euler circle gets its meaning by some social convention(s).

But one problem here is that the processes that apply to an image of a Euler circle are not the same as the processes that apply to the Euler circle in virtue of its conventionally determined meaning. Grant the internalist the point that the image of a Euler circle gets in my head because of some causal linkage with external Euler circles. The inferences that I make that involve Euler circles depend upon the conventions governing the properties and uses of Euler circles, something that the image cannot provide. So there are also mental representations of the conventions governing the properties and uses of Euler circles, which get in my head because of some causal linkage (asymmetric dependence say) with the outside world. But it is not the causal linkage that determines the content of these representations: the content of each of these representations *is* the convention. So unless there are mental representations with conventional content, there can be no cognitive processing of Euler circles.

This exchange between Clark and A&A illustrates their attempt to show that internal vehicles of cognition have intrinsic content, but that external vehicles don't. Given this stipulation, internal processing of vehicles with intrinsic content will always be cognitive. External processing of vehicles with conventionally derived content will not. The external vehicles and processes just don't exhibit the mark of the cognitive, but A&A's mark of the cognitive is seriously wanting.

6. Response to the Intrinsic Content Condition

Let's start with a simple example. A&A think that the image of a dog, *qua* mental representation in the brain, gets its content according to some naturalistic theory of content such as asymmetric dependence. The word 'dog' by contrast gets its content because of social conventions governing linguistic meaning.

There is a problem for A&A here. A&A say that internal cognitive processes cannot apply to conventional representations, only to images of them. However, as established above, the image of the Euler circle does not have the same content as the Euler circle. This is because the meaning of the Euler circle is dependent upon social conventions, but the content of the image is not determined by social conventions—A&A do not allow that images could have conventionally determined contents. But, as we saw above, internal processes don't apply to the image of an Euler circle in the same way that we directly manipulate the external Euler circle. Therefore, we cannot carry out the same operations on Euler circles in the head that we can by directly manipulating them if we are guided by A&A's stipulation. By A&A's own stipulation they are restricting the cognitive operations that can be carried out on Euler circles—and by extension all representations with socially determined content. If A&A were to concede that cognitive processes can involve vehicles with socially determined content, whether or not they are in the head, then their objection would become otiose.

By A&A's reasoning, cognition that involves representations with intrinsic content would be quite limited; but clearly we aren't so limited. This is because we have

developed external representational schemes and methods for manipulating them. It's odd, if not blindly stubborn, to think that such a large part of our resources required for completing cognitive tasks is not actually a part of our cognitive economy, simply because it involves the manipulation of vehicles with conventionally determined content.

There is another problem lurking here for A&A. They might strengthen the stipulation by insisting that the contents of all conventionally determined representations are derived from internal representations with intrinsic content. This would apply to sentences of natural language, stop signs and Euler circles (this is the strong version, held by, e.g., Fodor). But this is problematic because the meaning of an Euler circle is not equivalent to an image in someone's head. Take a different example: the mental image of an aardvark is not the sole constituent of my concept of an aardvark. In fact most of the content of my concept of an aardvark will have been fixed by the conventional methods that A&A find anathema. Concepts go beyond what asymmetric dependence or function of indication can offer. Clearly, any thinking I do by applying my concept of aardvark will be cognitive even though the concept is largely conventionally determined. According to the stricter interpretation of A&A's stipulation, the only genuine thinking that could be done would involve the image!

An obvious difference between an Euler circle and the image of an Euler circle is that images do not have conceptual contents. But then A&A may be forced to admit that representations with intrinsic content don't have the kind of contents that would figure in cognitive processes involving concepts. The image of a dog is not a symbol that can function as a constituent of a sentence in the language of thought. The word 'dog', by contrast, can be a constituent of sentences of English. If the only representations in the head with intrinsic content are imagistic, then intracranial cognition will be of an extremely basic kind. Of course integrationists think that it is precisely for these reasons that external symbol systems such as language and mathematics are required to transform our basic cognitive capacities.

Either A&A's intrinsic mental representations will be too rich—too similar to conventional representations—such that the intrinsic-conventional distinction becomes vacuous or they will be too meager, in which case they won't be of much use in completing cognitive tasks. Either way the intrinsic-derived distinction looks unhelpful.

A&A are left with a dilemma: either intrinsic representations and conventional representations have different meanings—as in the Euler circle case—in virtue of which limitations will have to be placed on the processing that we are capable of; or intrinsic representations and conventional representations have the same contents—either because they converge on the same meanings or one is dependent upon the other—in virtue of which there is no interesting distinction to be had for the cognitive case.

If, on the other hand, A&A are making a simple distinction on the grounds that images of natural objects and images of artifacts are about different kinds of things, then the insight is not of great importance to the integrationist. But if we took A&A's

money as good—that cognition involves vehicles with at least some intrinsic content—could this in any way pose a threat to integrationists? It is quite possible that cognition involves both external representations with “conventional” content and perceptual representations with “intrinsic” content. They might suppose that an internal visual representation of the external object is required for the manipulation of that object. But this would just be to say that the entire cognitive process involves some vehicles with intrinsic content and some with conventional content. If things turned out this way it would not harm the integrationist approach at all. The integrationist takes a cognitive practice—the manipulation of an external vehicle—to exhibit the mark of the cognitive because it is integrated with internal processes and it is aimed at the completion of a cognitive task.

As we saw above, once we understand the manipulation thesis properly then we know that the coupling constitution fallacy is aimed at the wrong target. When we understand a process as composed of internal and external manipulations, $X + Y$, then the combination of those manipulations, Z , can contain both internal vehicles with intrinsic contents and external vehicles with conventional contents. Therefore, in whatever way you pose the intrinsic content condition it is no worry for integrationists.

7. Extended Cognitive Science is no Science at all

The third attack involves the worry that there is little chance of finding any causal regularities covering both the brain and artifacts, such that we might construct a “brain-tool science.” Tools, as such, do not constitute a natural kind—e.g., DVD players and FM radios process information differently from one another. It follows that a brain-tool science would have to cover a motley of processes and thus could not be the basis of a genuine scientific enterprise.

Take C&C’s Otto example. Otto receives input from the environment, that there is an exhibition on at the MOMA, and he then retrieves the location of MOMA from his extended memory system, which causes him, as behavioral output, to go to 53rd street. Inga receives the same input from the environment, retrieves the information from a part of her brain, which then causes her to go to 53rd street.

C&C say that in the case of Otto and Inga there is a sufficient *functional similarity* between Otto’s use of his notebook and Inga’s recall from biological memory, that we are inclined to say that Otto has beliefs. I take this to mean the following: the functional role that Otto’s retrieval of information about the location of MOMA plays in his going to 53rd street, is sufficiently similar to the functional role that the pattern of activation in a part of Inga’s brain plays in her going to 53rd street, that we are inclined to say that Otto has beliefs.

As such the physical implementation of the functional role is irrelevant to the functional level of description. It is a matter of irrelevant implementational detail that Otto’s behavior is caused by his using his notebook and that Inga’s is caused by a pattern of activation in her brain. A&A’s problem is that Otto could make use

of a variety of media to recall the memory that MOMA is on 53rd street, e.g., notebooks, handheld PCs, CDs, etc. These media do not all work in the same way, which is supposed to lead to a motley of processes and vehicles and no genuine cognitive kinds—hence no real science.

However, I think that the worry trades upon an ambiguity between media and vehicle. For Otto it is the written sentence “MOMA is on 53rd street” that is the vehicle and this is distinct from the medium in which it is implemented—a page of a notebook, or the screen of a PDA. The contents of Otto and Inga’s beliefs are the same, even if they are accessing different vehicles to retrieve that belief content. Different media may prove to have different properties that affect access time to the contents—e.g., it might take Otto longer to retrieve the belief content than Inga, but that is a different issue.

A related worry is raised by Rupert (2004, pp. 407–415), who thinks that the Otto model of an extended memory system is not very plausible for understanding the case of the conversation-facilitating role of memory. He claims that in any conversation external vehicles, such as written language in notebooks, are not likely to play much of a role, whereas “internal storage” looks irreplaceable. The case of the “extended conversationalist” would involve the continuous writing down of what has just been said and referring back to it. Rupert claims that conversations are not like this; they depend upon the role of internal working memory. The conclusion Rupert draws is that we should reject the claim that the external processes in the Otto case could be constitutive of working memory, because the nature of the internal processes that make up working memory in the conversational case are so *dissimilar* to the processes in the Otto case.

The internalist case here needs to be not that internal and external processes differ, but that they differ in *important* and *relevant* respects which militate against the claim that the latter could be cognitive. A&A think that the important and relevant respect in which they differ is that internal processes involve intrinsic content and external ones don’t. It is less clear what the important and relevant differences are for Rupert.

This internalist argument, I suggest, only works if we take the parity principle, as a simple case of similarity, as a primary reason in favor of cognitive integration. However, if we do not consider the parity principle as a primary reason in favor of cognitive integration, then the dissimilarity objection has no force.⁴ A proper understanding of the manipulation thesis is required to deflect this internalist objection, as we’ll see below.

8. The Integrationist Response to the “Extended Cognitive Science is no Science at all” Objection

A&A and Rupert’s arguments trade on the claims that internal and external information processing are different, and that there is too much variety in the kinds of external information processing. This is because A&A and Rupert think that if they can show that internal vehicles and processes are different from external vehicles

and processes, then it follows that external vehicles and processes are not cognitive. I think that cognitive integrationists should accept the antecedent but deny that the consequent follows.

They should accept the first claim because it is a mistake to think of cognitive integration as just externalizing what is already in the head. It is true that the parity principle is apt to lead us in this direction: “They [C&C] contend that the active causal processes that extend into the environment *are just like the ones found in intracranial cognition*” (A&A, 2001, p. 56, original italics). Rupert (2004) makes precisely the same mistake: he clearly thinks that the downfall of cognitive integration is its attempt to show that external vehicles and processes are like internal ones.

My strategy is to focus on a specific kind of cognitive state, memory, and here the thrust of the discussion is twofold: I argue that the external portions of extended “memory” states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds; this quells *any temptation to argue for HEC* [hypothesis of extended cognition] *from brute analogy (namely, extended cognitive states are like wholly internal ones; therefore, they are of the same explanatory cognitive kind; therefore there are extended cognitive states)*. (p. 407, italics added)

However, cognitive integrationists who do not rely on the parity principle clearly avoid this criticism, because they take the manipulation thesis and the thesis of hybrid cognition to be the primary motivation for cognitive integration.

Instead, it [working memory] must be viewed as essentially *hybrid*, made up of two distinct components. In particular, the processes involved in working memory must be viewed as made up of both biological processes *and* processes of external manipulation of relevant information-bearing structures in the environment. (Rowlands, 1999, p. 147)

Remembering, on this view, involves exploiting internal, bodily, and environmental resources in order to produce some sort of action, often social in nature. (Wilson, 2004, p. 191)

There is nothing in the definition of working memory that Rupert chooses to work with that precludes the integrationist account. Working memory is part of “an integrated system for holding and manipulating information during the performance of complex cognitive tasks” (Baddeley, 2000, p. 78): it is precisely because internal and external components are integrated that they allow information to be available for the completion of cognitive tasks. But Rupert thinks that because he can show that internal memory and external memory function differently, he has undermined the whole cognitive integrationist project.

However, this can only be the case if cognitive integrationists proceed by showing that external processes involved in remembering get to be counted as *memory* because they are sufficiently similar to internal processes involved in remembering. The argument fails if we jettison the flawed version of the parity principle. Rowlands’ and Wilson’s accounts of the integration of internal and external memory fall quite happily under Rupert’s chosen definition of working memory.

If that line of attack is blocked, then Rupert could take each case on its merits. This he does with some success in the case of the role of memory in conversation. It seems likely that we follow conversations by drawing heavily on internal resources, but there are other cases of memory where we draw heavily on external resources. Assuming agreement with those psychologists cited by Rupert, why does it follow that all of memory is like the case of following a conversation? If all memory is internal in the way Rupert describes, then why is it easy for us to follow a conversation, but not so easy to recall the relevant detail at a later date? Why is it difficult to remember long strings of mathematical equations? Why do we need maps? We may not use many environmental cues in following a conversation, but we definitely use environmental cues in finding our way about. In these cases we don't need an explicit model in the head: sometimes the environment serves as its own best representation, or the representation in question is already in the environment—e.g., a map.

It is important to cognitive integration that external manipulations do something different to brain processes. In the head, there are connectionist vehicles and processes over them. There isn't anything in the environment that looks like connectionist vehicles and processes over them. There are symbols, such as diagrams and linguistically structured vehicles and their manipulation is different from the manipulation of connectionist vehicles—just think of the Euler circle case.

Otto's use of his notebook is cognitive because he manipulates the vehicles (sentences) in his book to complete a cognitive task. Inga completes the cognitive task using only her biological resources. Otto's manipulation of external vehicles is not cognitive because it is similar to Inga's biological memory, but because Otto and his notebook constitute "an integrated system for holding and manipulating information during the performance of complex cognitive tasks" (Baddeley, 2000, p. 78).

This move requires us to take seriously the complementarity of the internal and external cognitive resources necessary for their integration. There is a complementarity between what the biological brain can do and what the environment provides, such that inner processes and vehicles and outer processes and vehicles work together to complete a cognitive task (Clark, 2001, 2003; Sutton, 2006). We must also take into account the transformatory impact this integration has on our cognitive capacities, both in the here and now and during cognitive development. On this view the external cognitive environment transforms what the individual can do cognitively, both synchronically and diachronically (Menary, *in press*; Vygotsky, 1978; Wertsch, 1985).

So the processing of internal and external vehicles, whilst different, is complementary, and as integrationists we should try to show how they are complementary. However, it does not follow from this that external manipulations are a motley of processes. For example, there is an interesting difference between the Tetris case and the Otto case. In the Tetris case an epistemic action is performed in the place of an internal representation. In the case of Otto an external representation is used instead of an internal one. But Otto must have learnt to manipulate the notebook as a repository of representations. So Otto's bodily processes and

his manipulation of the representations in the notebook constitute his act of remembering. Similarly Tetris players learn to directly manipulate the differently shaped blocks, rather than rotate images of them in their heads. There is nothing particularly “motley” here; just two different kinds of manipulation that both allow for the completion of cognitive tasks.

A&A and Rupert miss the point here: certainly DVDs process “information” differently from brains, but it doesn’t follow from this that cognitive integration is false. When we manipulate external vehicles we aren’t doing what DVD players do! There are, as pointed out at the beginning of the paper, various types of manipulation:

1. Biological cases of coupling such as extended phenotypes and animate vision (biological coupling).
2. Using the environment as its own representation, obviating the need for internal representations—as in Tetris (epistemic actions).
3. And most importantly, the manipulation of external representational and notational systems according to certain normative practices—as in mathematics (cognitive practices).

There is nothing motley about these three classes of manipulation: what unifies them is that they *are* cases of the manipulation of external vehicles. Cognitive integrationists need to make these distinct cases clear, but once they are clear the internalists’ arguments fail.

9. Conclusion

Cognitive integration makes sense when we understand it as not just externalizing what is already in the head. The manipulation of external vehicles is importantly different from the manipulation of internal vehicles and their integration is the unit of cognitive analysis. We are not just coupling artifacts to pre-existing cognitive agents; the organism becomes a cognitive agent by being coupled to the external environment. Explaining this integration of the internal and external involves both a dynamical account of the reciprocal causal interaction between internal and external vehicles and processes, and an account of how we learn to manipulate external vehicles in accordance with relevant cognitive norms. Hence, the extended mind hypothesis contributes to the wider project of integrating the internal and external because it focuses on the dynamical account of causal coupling. Cognitive integration goes further than the extended mind hypothesis, because it explains how external cognitive vehicles are manipulated in a wider context where cognitive practices allow us to complete cognitive tasks.

Acknowledgements

Thanks to John Sutton, Andy Clark, Mark Rowlands, Dan Hutto, Mike Wheeler and Tim Bayne, for various comments and conversations.

Notes

- [1] A general definition of a cognitive task can easily end up being unhelpfully vacuous. If we define the cognitive task as any task for the completion of which cognition is required, then almost every task will be a cognitive one. I think it is more helpful if we think of cognitive tasks as involving the exercise of particular cognitive capacities such as remembering a date, solving a problem, learning to drive, etc. These are tasks where the exercising of cognitive capacities is directly tied to their successful completion.
- [2] C&C (1998, p. 8) give a longer version as follows: 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.
- [3] Here's a way of invoking a parity principle which does not lead to distortion of the integrationist's aims, suggested to me by Tim Bayne: Otto's actions, perceptions, etc. are, in the relevant ways, as causally coupled to his notebook as Inga's actions, perceptions, etc. are causally coupled to her hippocampus (or whatever). Of course there are causal differences between Otto's situation and Inga's situation, but these differences are not relevant differences.
- [4] It is certainly true that the parity principle is prominent in the formulation of the extended mind (Clark & Chalmers, 1998), but I think that Clark relies upon it as an intuition pump rather than as a primary reason for the hypothesis of extended mind.

References

- Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, 14, 43–64.
- Adams, F., & Aizawa, K. (2006). Defending the bounds of cognition. In R. Menary (Ed.), *The extended mind* (pp. 45–48). Aldershot, England: Ashgate.
- Baddeley, A. (2000). Short-term and working memory. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 77–92). New York: Oxford University Press.
- Ballard, D. (1991). Animate vision. *Artificial Intelligence*, 48, 57–86.
- Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: MIT Press.
- Clark, A. (2001). Reasons, robots, and the extended mind. *Mind and Language*, 16, 121–145.
- Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. Oxford, England: Oxford University Press.
- Clark, A. (2005). Intrinsic content, active memory and the extended mind. *Analysis*, 65, 1–11.
- Clark, A. (2006). Memento's revenge: The extended mind extended. In R. Menary (Ed.), *The extended mind* (pp. 18–44). Aldershot, England: Ashgate.
- Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58, 7–19.
- Dawkins, R. (1982). *The extended phenotype*. Oxford, England: Oxford University Press.
- Dretske, F. I. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- Fodor, J. A. (1990). *A theory of content*. Cambridge, MA: MIT Press.
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.
- Hurley, S. (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic actions. *Cognitive Science*, 18, 513–549.
- Menary, R. (in press). *Cognitive integration: Attacking the bounds of cognition*. Basingstoke, England: Palgrave.
- Millikan, R. G. (1984). *Language, thought and other biological categories*. Cambridge, MA: MIT Press.

- Millikan, R. G. (1993). *White Queen psychology and other essays for Alice*. Cambridge, MA: MIT Press.
- Millikan, R. G. (2004). *Varieties of meaning*. Cambridge, MA: MIT Press.
- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24, 939–973.
- Rowlands, M. (1999). *The body in mind: Understanding cognitive processes*. Cambridge, England: Cambridge University Press.
- Rumelhart, D. E., & McClelland, J. L. (1986). Schemata and sequential thought. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing* (Vol. 2, pp. 7–57). Cambridge, MA: MIT Press.
- Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, 101, 389–428.
- Sutton, J. (2006). Exograms and interdisciplinarity: History, the extended mind, and the civilizing process. In R. Menary (Ed.), *The extended mind* (pp. 234–297). Aldershot, England: Ashgate.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Webb, B. (1994). Robotic experiments in cricket phonotaxis. In D. Cliff, P. Husbands, J.-A Meyer, & S. Wilson (Eds.), *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior* (pp. 45–54). Cambridge, MA: MIT Press.
- Wertsch, J. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.
- Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. Cambridge, MA: MIT Press.
- Wilson, R. A. (2004). *Boundaries of the mind: The individual in the fragile sciences—cognition*. Cambridge, England: Cambridge University Press.