Sharing the blame:

Implications of the hypothesis of extended cognition

for personal identity and ethics

BA Philosophy Dissertation, 2013

Jessica Swallow

*I hereby certify that, except where references show otherwise, all the material contained in the attached dissertation is entirely my own work.*

Introduction……………………………………………………………........................... p.3

Chapter 1: Extended cognition……………………………………....…………………… p.4

1.1 Extending the mind: cognitive processing………………...……...... p.4

1.2 Extending the mind: mental states………………………………........ p.5

1.3 Parity and glue and trust……………………………………………………. p.7

1.4 Noteworthy objections to HEC……………………………...……………. p.8

1.4.1 The mark of the cognitive………………………………… p.9

1.4.2 The coupling-constitution fallacy………………………… p.10

1.4.3 Explanatory power: embedded vs. extended………….. p.11

1.5 Neuroethics………………………………………………………….............. p.12

1.5.1 Ethical parity………………………………………………………. p.13

1.5.2 Neurological interventions………………………............... p.15

1.5.3 Memory manipulation……………………………………........ p.16

1.6 Where does this leave us?..................................................... p.18

Chapter 2: Extended personal identity…………………………………………………. p.20

2.1 Wherein does personal identity consist?................................. p.21

2.1.1 The psychological criterion……………………………………. p.21

2.1.2 The biological criterion…………………………………………. p.22

2.1.3 The non-reductionist view: souls and egos…………….. p.23

2.1.4 Narrative identity…………………………………………………. p.24

2.1.5 Empirical evidence……………………………………………….. p.24

2.2 Personal identity and HEC………………………………………………….. p.26

2.2.1 Extended personal identity……………………………………. p.27

2.2.2 Ethical implications of extended personal identity……. p.28

Conclusion………………………………………………………………………………………… p.31

*Abstract: The hypothesis of extended cognition supposes that internal and external vehicles of cognition should be understood as being on a cognitive par; I propose that this requires that these vehicles should be treated as being on an ethical par. Further, I propose that the hypothesis entails extended personal identity, which enables us to make claims about the possibility of distributed and extended moral responsibility.*

**Introduction**

The hypothesis of extended cognition (or HEC, herein) has received a vast amount of attention in the philosophical literature. However, it is my belief that it has wide-ranging implications, many of which have been neglected in these discussions. In particular, as this paper demonstrates, I believe that HEC has both direct ethical implications, and also some which follow from the possibility of extended personal identity, which I also advocate.

The debate has reached something of a stalemate, with pro- and contra-HEC theorists resting on more or less liberal or chauvinistic functionalist intuitions (also known as the “Rowlands deadlock”; see Wheeler, 2010). I don’t propose to offer a way out of the impasse, only to expose some of the implications of HEC should it turn out to be true.

In this paper, I will begin in chapter 1 by setting up HEC as a plausible theory, defending it against some of its well-known criticisms; in section 1.6, I discuss some of the implications that it has for neuroethics. In chapter 2, I move to extended personal identity; in section 2.2, I propose that HEC entails extended personal identity, because persons are constituted by narrative memory, which, in most cases, is (partially) constituted by collective and extended processes. In subsection 2.2.2, I reach the ethical implications of extended and collective personal identity, concluding that moral accountability ought to be spread across this new, widely construed, person (to include social groups and notebooks), rather than focusing solely on the individual.

**Chapter 1: Extended cognition**

HEC was the name given to a radical view proposed by Clark and Chalmers (C&C herein) in their paper, ‘The Extended Mind’ (2010). Although similar theses have been defended under other names, including ‘environmentalism’ (Rowlands, 1999), ‘locational externalism’ (Wilson, 2004), and ‘content externalism’ (Burge, 1979; Putnam, 1975), in this paper I will work from C&C’s construal of the hypothesis. In short, their claim is that mental states aren’t constituted solely by brain states and that, in certain specified cases, the mind extends beyond the skin and skull and into the agent’s environment.

1.1 Extending the mind: cognitive processing

The argument for such a counterintuitive position as HEC is essentially functionalist. In the philosophy of mind, functionalism is a doctrine that supposes that what makes something a mental state is not its internal constitution, but the way in which it functions, or the role that it plays, in the system of which it is a part. C&C propose that agents frequently alter the world in order to aid and augment cognitive processes (known as “epistemic action” (Kirsh & Maglio, cited in C&C, 2010:28)), and that the human agent plus the use of an external entity in epistemic action therefore form a “coupled system” (C&C, 2010:29), which is a cognitive system in its own right. Just like in ordinary, internal cognition, all elements of the system play an active causal role, and they all jointly govern behaviour, with the external features being “just as causally relevant as typical internal features of the brain” (C&C, 2010:30). Hence, in line with functionalism, both these internal and extended systems may be correctly labelled ‘cognitive’.

To illustrate, consider that, when I write an essay, I don’t necessarily know how it is going to look when it is finished; it is only by placing my thoughts in some external medium (my laptop) that I am able to move them around, play with them, reflect on them, and form them fully. This is probably the case for most people. My laptop, so HEC says, is therefore involved in the cognitive process that produces my essay. Another example, from C&C, is the characterisation of my choice of words in Scrabble as the outcome of an extended cognitive process involving the rearrangement of tiles on my tray (C&C, 2010:30). Although this process could be explained in terms of internal processes and a stream of input and output thoughts and behaviours, the explanation would be “needlessly complex” (C&C, 2010:30), and if a similar process were to go on in the head, we would feel no need to characterise it in this way – we would surely just call the whole process ‘cognitive’. In short, “the rearrangement of tiles on the tray is not part of action; it is part of *thought*” (C&C, 2010:30).

There might be some initial resistance to the idea of extended cognitive processes. One issue could be that it would seem to follow that *consciousness* (insofar as it is a cognitive process, of sorts) extends beyond the skull: how could it be that my phenomenal experience of the world is extended? However, not every cognitive process is a conscious one (for example, the retrieval of memories, certain linguistic processes, or skill acquisition), and Clark concludes that extended vehicles of *conscious* experience aren’t poised for use in the same way that internal vehicles are, hence they aren’t functionally equivalent, hence consciousness isn’t an extended process. The debate about extended consciousness is an interesting one, but I leave it here, rebutted by Clark, as it is contentious and by no means required for a general acceptance of HEC (see Clark, 2009; Noë, 2007).

Another source of objection might be in the idea that the brain, unlike the external environment, is *portable*. It therefore forms the package of resources that we bring to bear when confronted with a cognitive task, regardless of the local environment: the problem with coupled systems being “that they are too easily *decoupled*” (C&C, 2010:31). There does seem to be something to this objection: the fact that the external environment is only contingently available, whereas the brain and body are present for all of a particular agent’s cognitive undertakings, might suggest that the brain and body are interesting cognitive resources in their own right, worthy of consideration as a ‘cognitive kind’ all of their own.

However, it’s conceivable that we could have various contingently available modules that could be plugged into our brain, like an external memory drive, for example. The fact that the external memory drive is only contingently plugged in surely does not rule out the possibility of its being included in the cognitive operations of the mind. The portability intuition, rather than dealing a damning blow to HEC, actually helps us to *refine* it: not just any old coupling to the environment will do; it must be *reliable* coupling. It might so happen that the most reliable coupling of resources occurs in the brain, but there is no good reason to suppose that reliable enough coupling cannot go on between an agent and her environment (for example, she might carry around her trusty nautical slide rule at all times). It isn’t even enough to say that the nautical slide rule might get lost or damaged, since this is also the case with brains.

1.2 Extending the mind: mental states

So we have started to see how some cognitive *processing* might take place in the environment, but HEC goes one step further to say that the same might be true of certain other mental *states* (including non-occurrent, or dispositional, beliefs). The idea that a belief could be stored in an external medium might sound odd, but there are some convincing arguments for its acceptance.

C&C introduce us to Otto (2010:33-37). Sadly, Otto suffers from Alzheimer’s disease, but he functions normally thanks to his handy notebook. He carries it around with him at all times and it contains everything that one might hope to be stored in one’s ordinary memory, including his long-term memories, such as where he grew up and what his favourite game was as a child; information, like who he banks with and when his children were born; long-term desires, like his dream holiday destination; etc. Most importantly for C&C’s story, it contains the address of the Museum of Modern Art (MoMA). Otto finds out that there is an exhibition on at MoMA, and he decides to go and see it. He looks the address up in his notebook (53rd Street), and makes his way there.

Now meet Inga. She doesn’t have Alzheimer’s and is able to store beliefs in her head. She, like Otto, learns that there is an exhibition on at MoMA and she decides that she would like to go. She remembers that MoMA is on 53rd Street and makes her way there.

C&C claim that all the relevant aspects in the cases of Otto and Inga are “entirely analogous” (2010:33). Since Otto’s notebook and Inga’s memory perform the same functional role in guiding behaviour, and we would treat Inga’s consultation of her memory as a cognitive processes, the proposal is that we ought to treat the consultation of *both*, respectively, as a cognitive *process*.

Further, if we can say that Inga had the standing belief that MoMA was on 53rd Street *before* she consulted her memory – a point that isn’t usually contested – then we also ought to be able to say that Otto had the same kind of belief *before he consulted his notebook.* That is to say, Otto’s notebook, in this case, functions *exactly* *like* Inga’s memory, so the information contained within it is *exactly like* Inga’s belief.[[1]](#footnote-1) The alternative, in Otto’s case, would be to concede that he has no belief at all before he consults his notebook, and that, “at best, he believes that the museum is located at the address in his notebook” (C&C, 2010:34). But this doesn’t seem to do his notebook justice: if we follow Otto around for a while, we will see that he is constantly using it as a matter of course; it is central to his actions in many different contexts, “in the [same] way that an ordinary memory is central in an ordinary life” (C&C, 2010:34).

We might say that the problem is that Otto’s beliefs disappear when he puts his notebook away, but so do Inga’s when they aren’t at the forefront of her mind. It would miss the bigger picture with Inga to say that her beliefs no longer exist when she isn’t using them, and it would be the same for Otto. So it must either be that Otto, like Inga, has an *occurrent desire* to go to the museum and a *dispositional belief* that the museum is on 53rd street, or that he has an occurrent desire to go to the museum, a dispositional belief that the museum is at whatever location his notebook says, and an accessible fact that the notebook says that the museum is on 53rd street. The latter would be over-complicated; it is to take “one step too many” (C&C, 2010:34), and it would be pointlessly complex to explain Inga’s actions in terms of her beliefs about her memory for similar reasons, so why not think the same for Otto?

In addition to extending the mind into inanimate objects, it seems likely that *other people* might constitute parts of our cognitive system. Groups of people are often organised in such a way that they perform the functions of minds: for instance, juries might be an example of distributed problem-solving, while scientific communities also learn, work, and establish facts as a group (see Latour, 1987; Fleck, 1986). In both cases, individuals stand in a dynamic relation to one another, operating as one to reach some final end. It is not necessary to delve any further into the topic of socially extended cognition at this stage; I merely hoped to introduce the reader to this possibility, as it will be relevant for some of the considerations that follow (see Sutton, 2010b; Hutchins, 2001).

1.3 Parity and glue and trust

The implication of functional similarity is made explicit through the Parity Principle:

“If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process” (Clark, 2010a:44).

The principle means to say that if a particular state or process would be called cognitive if it went on in the head, then there is no good reason other than internalist chauvinism to suppose that the same state or process, if it goes on *outside* the head, is *not* cognitive. In this case, the external resource is somehow *coupled* to the agent so that it functions as part of her cognitive apparatus.

There are, however, only certain kinds of coupling that are relevant for HEC. There are some strange implications if we buy into HEC based on parity alone. For example, I use Google on my iPhone quite a lot; I can look up the name of the actor who plays Leonard in *Memento* (Guy Pearce); I can find out how many miles there are between Exeter and Prague (962). If I knew these facts without having to look them up, we would probably award these ‘beliefs’ cognitive status. So, on parity, we should say that I *believe* that Guy Pearce plays Leonard in *Memento* even before I look it up on my iPhone. So it seems that I believe everything on the Internet (Allen-Hermanson, 2012:2). Alternatively, perhaps Inga is feeling spiteful and has changed the address of MoMA in Otto’s notebook so that it reads “51st Street”. Does Otto then believe that the exhibition is on 51st Street? The issue at hand is the “problem of cognitive bloat” (Allen-Hermanson, 2012): if C&C opt for the simplistic view that anything that is causally connected to a cognitive process is thereby constitutive of it, then there is “the threat of cognition bleeding into everything” (Adams and Aizawa, 2001:57).

To avoid this problem, C&C stipulate certain (“glue and trust”) conditions, the satisfaction of which guarantees that any candidate for cognition is coupled to the agent in the right kind of way. They are as follows:

1. The resource is reliable and typically invoked.
2. Any information in the resource is more-or-less automatically endorsed.
3. The information in the resource is easily accessible when required.
4. The information in the resource has been consciously endorsed at some point in the past (C&C, 2010:38).

The fourth condition as it was proposed in the original paper is, however, contentious, because it would seem to entail that it would be incorrect to call a number of ordinary, internal representations ‘beliefs’, such as those resulting from subliminal perception or memory tampering (Rupert, 2004). For present purposes, I will be taking the requirements for relevant coupling to be the more generally accepted conditions 1-3.

So let’s say that Otto’s notebook is ‘coupled’ to Otto inasmuch as he uses it in most situations, he rarely doubts its contents, and he can easily find the information he requires within it, perhaps through the use of some elaborate colour-coding system. For C&C, Otto’s notebook satisfies all of the relevant criteria for being considered a literal part of Otto’s mind. This is the point at which the ethical considerations with which I began this paper become apparent: if Otto’s notebook is considered a proper part of Otto’s mind, then it ought to be treated in a particular way. It is clear that if Inga took a match to Otto’s notebook, she would be committing some *wrong* that would, conversely, not be applicable were she to throw my calculator (which *isn’t* relevantly coupled to me very often, since I only use it now and again, and it often breaks) in a river. I will return to this idea towards the end of this chapter, once I have done a little more work to convince the reader of the plausibility of HEC.

1.4 Noteworthy objections to HEC

There are many noteworthy objections to the parity principle and the glue and trust conditions. In the following I will focus on some from Rupert (2004) and Adams and Aizawa (A&A herein; 2010, 2008, 2001), though there are others worthy of consideration, including Sprevak’s concern about the relevant grain of functional equivalence (2009), and Sutton’s preference for the ‘complementarity principle’ as a better basis for HEC than parity (2010a). I will spend a moment addressing Rupert and A&A, as versions of their criticisms are usually taken to be some of the most central challenges to HEC. I do not hope, and nor have I tried, to refute all of the possible criticisms that have been brought against HEC. Rather, I hope to set it up as a theory that is plausible enough to warrant investigation into some of its implications, which I will come to later.

*1.4.1 The mark of the cognitive*

Rupert (2004) raises concerns about the plausibility of HEC due to the profound differences that distinguish inner and outer contributions to memory. Therefore, he argues, they ought to be treated as distinct kinds, rather than as one overarching ‘memory’ system (see also A&A, 2010).

Rupert considers the display of negative transfer and the generation effect in normal, biological memory systems. In paired-associates experiments, subjects learn assigned associations between pairs of stimulus-items, and their ability to recall these associations is tested in different ways. The associations might be between a list (A) of men’s names and another list (B) of their female spouses. Subjects display what is called ‘negative transfer’ if they are asked to memorise a different set of female names (C), which are associated with the original list A. They learn the A-C associations significantly more slowly than the A-B associations because there is “an interference of the old associations with the learning of the new” (Rupert, 2004:31; see also Anderson, 2000).

Rupert correctly points out that there is no reason to suppose that subjects would display negative transfer if they relied on external stores. If they have the A-B and A-C lists written down in a notebook, it should be no harder to ‘recall’ the A-C list than the A-B. In fact, there will be no interference whatsoever between the two lists; this is vastly unlike biological memory, where individual memory items stand in a dynamic relation to one another.

Further, in similar experiments, subjects are asked to come up with their own pairs of associations. They are shown to perform significantly better than others who have been given the associations to be memorised (Bobrow & Bower, 1969). This is known as the ‘generation effect’. Again, if experimenters gave notebooks containing the correct associations to one group of subjects, and gave blank notebooks to another group and asked them to come up with their own, there is no reason to suppose that there would be any difference in performance between the two when asked to ‘recall’ the associations. Both groups would be able to access the notebook equally well and give equally correct answers.

Rupert’s point here is that the negative transfer and the generation effect are both features of *internal*, but not *external* memory. Since there are disparities between the operational capabilities of internal and external memory devices, he argues, they are of different explanatory kinds.

If the parity principle were to be understood as a requirement of deep similarity between inner and outer processes, Rupert might have dealt a knockout blow to HEC. But this isn’t the case; it was simply intended by C&C to undermine the tendency to understand inner human processes as the standard against which any candidate for cognitive processes must be measured. Rather than requiring any potentially cognitive process to conform entirely to the features of other processes known to be cognitive, the principle should be understood as levelling the playing field between the inner and outer candidates for cognition, allowing any process resembling one that might intuitively be judged to be cognitive to be granted such a status.

So Rupert takes there to be certain features that uniquely belong to biological memory systems: the ‘mark of memory’ (or the ‘mark of the cognitive’, see A&A, 2001). However, Clark succinctly responds to this objection, positing that the generation effect (or negative transfer, or intrinsic content[[2]](#footnote-2)) may be a feature of internal systems but, “[i]n general, for some *X* to be part of the supervenience base of some *Y*, where that *Y* must… exhibit some property *Z*, there is no requirement *that Z be in addition a property of the putative part X*” (2010b:89). The flaw in Rupert’s objection is in assuming that all parts of the memory (or cognitive) system must exhibit the generation effect when, if it is in fact a requirement of the system, it is enough that the *whole system* exhibits it in its proper, unified operation.

Further still, certain areas of biological memory systems do not exhibit the generation effect – the recall of answers to arithmetic problems, for example (McNamara & Healey, 1995) – and yet we would surely not deny their place in memory. If we are prepared to afford leniency to internal systems in terms of their exhibition of ‘marks’, then there seems to be no good reason not to do the same for extended systems.

*1.4.2 The coupling-constitution fallacy*

Suppose that we can successfully rebut Rupert’s accusation regarding the mark of the cognitive. So we can now allow that external resources are candidate parts of a cognitive system by virtue of being relevantly coupled to an agent, even though they don’t behave exactly like internal resources. But, even so, “to move from the causal coupling of some object or process to some cognitive agent, to the conclusion that the object or process is part of the cognitive agent, or part of the agent’s cognitive processing” (Clark, 2010b:81) would be fallacious. Hence, A&A accuse HEC-theorists of committing the “coupling constitution fallacy” (A&A, 2010:67).

However, their criticism, as they formulate it, lacks weight. They mockingly ask “Why did the pencil think that 2+2=4?” and give “Clark’s answer: Because it was coupled to the mathematician” (A&A, 2010a:67). Yet HEC-theorists make no claim that relevant coupling means that the external resource itself becomes cognitive; instead, satisfaction of parity and glue and trust ensures that the object is *part of a cognitive system*. The pencil isn’t cognitive, but forms a proper part of the wider system that is the mathematician’s cognitive apparatus.

Even if the objection isn’t so obtuse as to attack such a straw man, it doesn’t hold. It seems that A&A believe that “some objects or processes, *in virtue of their own nature*… are… *candidate parts* (for inclusion in a cognitive process), whereas other objects or processes, still in virtue of their own nature, are not” (Clark, 2010b:84). However, the nature and boundaries of cognitive systems are exactly what is in question. To make a distinction between causally and constitutively related resources requires some prior assumption about the boundary of the system in question. The coupling-constitution fallacy therefore only applies when the boundaries are known (as far as they can be). Appeal to common sense division of cognitive and non-cognitive kinds is simply question begging in this case (Ross & Ladyman, 2010:158).

*1.4.3 Explanatory power: embedded vs. extended*

Many theorists agree that cognitive processes depend heavily, and intimately, on external props and devices, and on the structure of the external environment in which cognition takes place. Some take the ‘embodied’ approach to cognition (including Lakoff & Johnson, 1999), which appeals to the idea that it depends upon aspects of the agent’s *body*, as well the brain, and that, without the involvement of the body in both sensing and acting, “thoughts would be empty, and mental affairs would not exhibit the characteristics and properties they do” (Wilson & Foglia, 2011:§1).

Another alternative is the ‘embedded’ approach (the hypothesis of embedded cognition, or HEMC, herein). This draws on the idea that cognition is deeply dependent on the natural and social environment of the agent, and places particular emphasis on the way in which the agent offloads some cognitive work onto external media, which deeply and dynamically supports cognition (see Hutchins, 1995; Suchman, 1987).

HEC and HEMC are competing theories of cognitive phenomena, where HEMC concedes that cognition is (partially) *dependent on* the environment, while HEC claims that the environment (partially) *constitutes* cognition. Rupert, for example, suggests that both theories are right to emphasise the importance of external devices to cognition, and that we ought to accept whichever one provides superior explanations of phenomena to cognitive science. He believes that HEMC is the way to go, not only because of his aforementioned concerns, but also because HEC doesn’t provide a helpful taxonomy: the diversity of structures and processes present in an extended cognitive system would mean that we would be restricted to talking only about “generic” cognition (2004:37).

Similarly, A&A argue that it is the job of science to “carve nature at its joints” (2008:58); yet a “brain-tool science” (2010:73), which lumps together intracranial and extracranial goings-on, would form an “unruly collection of processes” (2010:73) that have little to nothing in common. There must be common features or regularities that tie together processes of a common kind, and it is unlikely that the causal arrangements enabling external processes to contribute to action will display any similarities to their internal counterparts. If this is the case then there can be no unified science of the extended mind. Better to have a unified science of the inner, and a separate, unified science of the outer.

However, I align myself with Clark in saying that it is dangerous to claim, from the armchair, that there are no regularities to be found (2010a:50); an aim of a science is arguably to establish taxonomies by its very working, not to decide on them in advance and work out what does and does not fit the description. If this is the case then the parity principle serves us well, preventing us from drawing boundaries based on intuitions and enabling us to establish them based on facts about the perceived phenomena.

Further, it isn’t obvious that unification of a system requires that all elements of that system behave according to the same laws, and it is quite possible, even, that the huge variety of *inner* goings on that A&A prefer to talk about will form such a motley of causal mechanisms that they won’t bear even a family resemblance. As discussed in section 1.4.1, even some less contentious internal elements of cognition don’t behave according to the same laws as each other; if we are going to allow this motley to be understood as a unified system, this criticism cannot be brought against HEC.

The question of exactly how useful HEC is as a framework for cognitive science has perhaps not yet been answered decisively (see Pöyhönen, forthcoming). This said, I am convinced enough of the truth of HEC to warrant further investigation into its implications, which might either be referred to or dismissed entirely should some future evidence sway the discussion in one direction or another. However, in the interests of concision, from here on in I will take the truth of HEC as a given, referring only to HEMC as appropriate.

1.5 Neuroethics

Neuroethics is concerned with the ethical questions that are bound up with the sciences of the mind. It has two main branches: the *ethics of neuroscience* and the *neuroscience of ethics* (Roskies, 2002). The former refers to the body of work that seeks to develop an ethical framework for neuroscientific enquiry; the latter refers to the impact of neuroscience on current ethical theory. Although the ethics of neuroscience partly concerns the ethics of neuroscientific *conduct* (including the ethics of withholding particular findings, etc), in the following I will focus on ethical concerns about the *application* of neuroscientific knowledge, including the use of psychopharmaceuticals and other such interventions into the mind.

Due to recent technological advances, neuroscientists are increasingly able to intervene in the brains of subjects – they are already able to alter personality traits and enhance cognitive capacities, and things like the ability to read minds or insert beliefs don’t seem like such distant possibilities anymore. Although perhaps a little disconcerting, this research is important because it promises to answer questions about how some of our most fundamentally human capacities work, such as our free will, and our ability to control our actions and desires.

Ethical questions surrounding these techniques tend to focus on whether interventions into the mind are ethically sound, or whether they threaten important elements of our autonomy. However, in light of HEC, it seems likely that interventions into the mind occur with great regularity, from reading people’s diaries to inputting your number to their phone. HEC thereby significantly widens the scope of neuroethics, and shifts its focus away from questions about *whether* interventions into the mind are ethically sound to ones about *which* interventions into the (widely construed) mind we ought to allow, and under what conditions. It will become clear why this is the case as we go on.

*1.5.1 Ethical parity*

Based on the parity principle, HEC says that the distinction between mind and world is essentially arbitrary, and that any process that would be called cognitive if it were internal should similarly be called cognitive if it goes on externally. Hence, extended cognitive systems exist, which replicate the functions of certain internal process, as well as greatly enhancing and improving them, in some cases. Following this, Levy proposes an ethical construal of the parity principle, the “ethical parity principle” (EPP), which, in its strongest form, asserts: “Since the mind extends into the external environment, alterations of external props used for thinking are (*ceteris paribus*) ethically on a par with alterations of the brain” (Levy, 2007a:61). In other words, our ethical responses to interventions into the cognitive environment ought to be consistent with our ethical responses to interventions into the brain.

However, it’s worth noting that HEC doesn’t need to be true in order for interventions into the cognitive environment to be significant; HEMC also takes cognition to be heavily dependent on environmental scaffolding, so acceptance of HEC is not a prerequisite for the equality of ethical judgements about these interventions. Levy also allows room for this by proposing a weaker version of EPP: if our reasons for finding alterations of the brain problematic are transferrable to alterations of the environment, then alterations of the brain are ethically on a par with alterations of the environment (Levy, 2007a:61). I will, for present purposes, continue to assume the truth of HEC, but the HEMC-theorist also ought to examine the reasons that one might have for finding some interventions into or alterations of the (narrowly construed) mind impermissible or problematic, and decide whether they apply equally strongly to actual or possible interventions into or alterations of the environmental scaffold that supports the mind. If they do, then one ought to hold that internal and external interventions are (*ceteris paribus*) equally problematic.[[3]](#footnote-3)

To illustrate, consider that we might replace Otto’s notebook with an iPad, which will be as much of a constant in his life as his notebook ever was. His cognitive performance would be greatly improved as he would be able to find information much faster by using the iPad’s search function, he would also be able to look things up on the Internet, and make use of the calculator and map ‘apps’, etc. We might like to claim that he is ‘cheating’, that the use of the iPad enhances his cognitive capacities beyond the normal human range, just as we might like to say that taking methylphenidate (Ritalin) to enhance cognitive performance is ‘cheating’ (see Harris & Chatterjee, 2009). Conversely, if the replacement of Otto’s notebook with an iPad is not ethically problematic, and raises no ethically significant questions at all, then we ought not regard analogous interventions into the brain (like taking methylphenidate) as ethically problematic. In other words, they cannot be deemed ‘wrong’ simply by virtue of the fact that they are interventions into the brain if the brain/world distinction has been sufficiently undermined by EPP (Levy, 2007a:61).

It is worth noting, however, that the *“ceteris paribus*” within EPP is doing rather a lot of work. The parity principle for HEC ensures that external resources are granted cognitive status if they do the same work as internal ones, but we ought to be wary of equating cognitive extension with ethical extension: the kinds of differences between internal and external resources and processes that HEC considers to be irrelevant might turn out to be “ethically relevant” (Levy, 2007a:61). For example, the fact that Otto’s iPad might be easily replaced (because he has properly backed it up), whereas Inga’s hippocampus cannot be, is not relevant to HEC: quite simply, while they both play the same kind of role, they are both parts of cognitive systems. However, if we consider this scenario from an ethical perspective, it seems to matter very much that Otto’s iPad is replaceable and Inga’s hippocampus isn’t: damaging Otto’s iPad seems less serious than damaging Inga’s hippocampus. So Otto’s iPad is only of ethical concern if he has *not* backed it up because, in that case, damage to it could be just as detrimental to him as damage to Inga’s hippocampus could be detrimental to her.

*1.5.2 Neurological interventions*

So acceptance of EPP (in either its weak or strong form) amounts to a rejection of the claim that “interventions into *the brain* are uniquely and distinctively interventions into the mental states that constitute our identities” (Levy, 2007b:7), which, in turn, underlies the intuition that neurological interventions (like direct brain stimulation, etc.) are fundamentally different from more traditional methods of altering mental states (like talking or behavioural therapy, or rational argument). Said intuition might have been defended on the grounds that traditional ways of changing minds involve the *assimilation* of new information into the mind of the agent, whereas new technologies *remain* “external and alien intrusions into the mind” (Levy, 2007b:8) (i.e. the mind doesn’t incorporate them into its workings but, rather, is directly changed by them). However, given HEC, these new technologies aren’t so different from the traditional methods: if there is, historically, a wide range of cases in which external manipulations have shaped human cognition, then the claim that psychopharmaceuticals, for example, are especially suspect, seems much less plausible.

Another possible objection to neurological intervention pertains specifically to the possibility of cognitive *enhancement*. As I mentioned earlier, it might be said that taking methylphenidate to boost cognitive performance would be ‘cheating’; this could be considered both in relation to other individuals, as it could undermine their genuine effort or talent, and in relation to the drug-user, as he fails to build the strength of character that practice and effort help to develop. However, taking vitamins or hiring a personal tutor, for example, aren’t usually thought of in such a negative light. This might be because methylphenidate targets the mind, whereas vitamins target the body and the personal tutor uses traditional methods of teaching, but, based on EPP, alterations to the body or external environment are arguably on an ethical par with alterations of the brain, since they all involve interventions into the mind, widely construed. In short, if our environmental manipulations frequently enhance the mind, then the claim that neuroenhancement of the kind at issue is *especially* problematic requires some defence.

So it appears that, as there seems to be within the broader fields of cognitive science and philosophy of mind, there may well be an undefended internalist privilege at play within traditional neuroethics. This may well be because of a lack of a persuasive alternative, but, given HEC, this is no longer the case. If neuroethicists want to establish that this neurological intervention is more or less problematic than that neurological intervention, they ought not rely on undefended internalist assumptions. This said, there might be some internal interventions that raise more, and more serious, worries than external (like the above example involving the removal of Inga’s hippocampus), but, in these cases, an argument for internalism is required as the “crucial first step” (Levy, 2007b:9).

Through HEC, it seems that interventions into the *brain* might not be so uniquely and distinctively ethically relevant: interventions into the cognitive environment, in at least some cases, might be on an ethical par. If this is the case, then it becomes less clear that neuroethics ought to be concerned with *whether* we ought to allow interventions into the mind, since this sort of question only makes sense if we have a choice about whether or not they occur. If there is parity between internal and external means of changing minds then it seems that many of the things that we routinely do might alter and enhance certain mental states and processes (be they our own, or those of others), so we must conclude that “we have no choice about whether to allow interventions into the mind” (Levy, 2007b:9). So the important question becomes one about *which* interventions are permissible.

*1.5.3 Memory manipulation*

To further demonstrate the kind of ethically relevant neurological interventions to which I refer, imagine that memory erasure, like the kind envisaged in films like *Eternal Sunshine of the Spotless Mind* (2004), is possible.[[4]](#footnote-4) Our memories are important to our sense of identity over time as they form our individual life-narrative; they help us to make sense of our actions, desires, and personalities, which are made significant because they are situated within the context of an unfolding life story. Just how important memory is to our identity over time will be discussed in the next chapter; for now it is enough that the reader has an intuitive agreement about the importance of memory to individuals, at least to some degree.

Memory erasure could certainly cause harm to *ourselves*. By remembering past experiences – both our successes and our mistakes – we are able to learn from them, to grow as individuals, and to increase our own self-knowledge. “Erase our memories, and we leave ourselves at the mercy of impulses of the moment” (Levy, 2007a:174); we become an illustration of Santayana’s dictum that “those who cannot remember the past are condemned to repeat it” (Santayana, 1905:285, cited in Meyer, 2008:77). However, manipulation of our own memory wouldn’t be unprecedented; it is common knowledge that we sometimes supress traumatic memories, or just forget what we had for breakfast, though these kinds of memory erasure are arguably involuntary, to some degree. But experiments have shown that we can even *voluntarily* forget things: subjects were instructed to forget items, like a list of words presented to them, and they were shown to have some success at the task (Whetstone & Cross, 1998). These results, however, aren’t as drastic as those displayed in *Eternal Sunshine*, or those that may be possible in the future through the use of neurological intervention technologies.

Harm that I may or may not cause to myself, however, isn’t so ethically interesting as the harm that I might cause to others, since I arguably have a right to harm myself, but not others. In line with Mill’s harm principle (2006), I suggest that each of us has the right to think and act as we will, as long as our actions don’t result in harm to (non-consenting) others. Application of the harm principle yields the result that I can only erase my memories when it harms no one, or when it harms only those who have consented to it (Levy, 2007a:178). This said, there may be cases in which the harm caused to others is relatively superficial in comparison with the kind of fundamental good that that same action might do for me, but, “[i]n general, we are allowed to cause relatively minor harms, if we have no other practical way of pursuing goals and projects central to our conception of the good life” (Levy, 2007a:179).

So the permissibility of memory erasure depends not on whether harm is done to others, but on the *extent* to which erasure of my own memory might cause harm. So how might erasure of my memories cause harm to someone else? As Levy suggests, we use other people as reference points from which we can check our own self-narrative. Many of us are familiar with the experience of misremembering something, and having our friends pick us up on it. Clearly, if I erase my memories of you, “I remove myself, permanently and irrevocably, as a reference point against which you can check your self-narrative” (Levy, 2007a:175).

More damaging, perhaps, would be the fact that I wouldn’t even *recognise* you. In Ralph Ellison’s novel, *The Invisible Man* (1995), the nameless protagonist tells of his “invisibility”, which is not a physical attribute, but a comment on the fact that other people refuse to acknowledge him. It transpires that the narrator is black, and that his experience of being invisible is the result of a ubiquitous, humiliating, and subtle form of racism. Taking the Hegelian conception of the “struggle for recognition” (Hegel, 1977), it can be claimed that “as a necessary condition for acquiring beliefs about our own beliefs and desires, we need the recognition of another person” (Honneth & Margalit, 2001:128; see also Redding, 2012). Hegel also proposed that our sense of self-*worth* depends upon recognition from others who we ourselves deem worthy of recognition in turn. Recognition from others is important to our sense of identity because, as Taylor (1995) points out, there is an extent to which our identity is *dialogical*; that is, the person who I am is partially formed through dialogue with others. Humans are cultural animals, our identity is bound up in the cultures with which we identify ourselves, hence it is important to us to be recognised as a member of that culture.

So lack of recognition can be damaging to our sense of identity – but why? Imagine a couple that have separated on bad terms. It might be expected that they would ignore each other if they were to bump into one another, but there is an important sense in which this *would* constitute recognition of some sort. In order to share the kind of contempt for each other that they do, they are required to recognise that they once shared a life together. If, on the other hand, one of them were to have his/her memories of the other erased, it would be as if the relationship had never been. The identity-constituting narrative that existed only out of the relationship and interaction of the pair ceases to exist meaningfully if one of them has no recollection of it.

Memory erasure, therefore, looks to be a potential source of harm not only to the subject of the erasure alone, but also to any individuals whom she has opted to erase from her own memory. Having established this idea within the bounds of the traditional internalist conception of the mind, it becomes clear that HEC complicates things. If the mind were confined to the skull, there would be a strong sense in which the individual ought to have sovereignty over her own mind. However, if the mind extends into the world, then it seems that altering one’s own mind might also alter the minds of others, hence it seems less likely that individuals ought to retain this sovereignty. Rather than the weaker sense in which recognition by other individuals might be important to me, it appears that there could well be a very real way in which, under certain conditions, they could be repositories for some of my memories. If this is the case, then the erasure of my memories of an ex-partner might cause him *direct* harm, rather than just denying him worth-conferring recognition.

Further, HEC not only alters the way in which we might deem hypothetical, sci-fi memory erasure to be permissible, it also seems to make this hypothetical situation into a here-and-now reality: taking a match to Otto’s notebook would *literally erase his memories.* HEC in combination with the harm principle therefore rules out a great many more actions as impermissible than the internalism-plus-harm-principle alternative.

1.6 Where does this leave us?

HEC ought to change the way in which we think about interventions into the mind (widely construed). It seems that we have two options: on the one hand, we accept that neurological interventions are permissible on a scale that we might have intuitively deemed unthinkable. For example, given that we frequently alter or intervene in people’s external memory stores, *ceteris paribus*, we ought not have any problem with analogous interventions into the *brain*. If we can replace Otto’s notebook with an iPad, then taking methylphenidate should be permissible; if I can input my number to your phone, then inserting similar beliefs directly into your brain should also be permissible.

On the other hand, the alternative would be to admit the significance of the resources that make up an individual’s cognitive environment, and *deny* the permissibility of certain manipulations, on the grounds that we would deny the permissibility of the analogous neurological intervention. Destroying Otto’s notebook would be, *ceteris paribus*, on an ethical par with destroying Inga’s hippocampus, and so would require a similar ethical judgement.

In either case, it appears that HEC is prescriptive of a particular kind of treatment of brains and cognitive resources. HEMC, too, draws attention to these kinds of concerns, but it has a little more legwork to do in order to make the jump to ethical parity; HEC makes the case for the inclusion of particular resources in the mind, and, since the mind is ethically relevant, so must the external resources be. HEMC would need to grade external resources according to the extent to which they scaffold the mind, and then qualify the claim that some interventions would be permissible, and some wouldn’t.

This aside, I hope to have demonstrated that HEC has clear implications for neuroethics: if it is the case that our minds *literally extend into the environment*, then we should be able to expect treatment of our relevant external props to be aligned with treatment of our minds, in accordance with our dignity as cognisers, and as persons.

**Chapter 2: Extended personal identity**

As I have discussed, HEC has certain direct ethical implications, including a loosening of the ethical distinction between neurological and environmental interventions, and an implication of equality of treatment of brains and the external devices that constitute mental states or processes. I believe that it has much further-reaching ethical implications, however, since it entails that *personal identity*, in at least some cases, is extended.

I wish to make it explicit from the start that I am not talking here about extended *persons*; this idea has had some attention in the literature, from C&C’s tentative proposal of it in their original essay (“Does the extended mind imply an extended self? It seems so.” (C&C, 2010:39)) to Olson’s staunch refutation of it (2011). Instead, I follow Wilson and Lenart (forthcoming) in proposing that, in at least some cases, we keep track of ourselves over time, maintaining a single, albeit evolving, identity and personality, *by virtue of the fact that we are exposed to persisting environmental cues*.

To be clear, I understand a ‘person’ to be a particular human to which certain biological and psychological states are attributable, and who can be understood as existing at a single point in time. ‘Personal identity’, then, is the property by which an individual is deemed to be *the same person* as she was yesterday, or last year, or ten years ago, etc, and it is necessarily a relation that exists *over time*. Further, it is usually thought of as being a *numerical* identity (or a one-to-one identity), meaning that there can be no two individuals who can be considered to be the same person.

All of this relates to what Schechtman calls the *reidentification question*: “what makes a person at time t2 the same person as a person at time t1”? (1996:1-2). She proposes a separation of the question of reidentification from that of *characterisation* within discussions of personal identity, the latter being “the question of which beliefs, values, desires, and other psychological features make someone the person she is” (1996:2). What I propose is that, in some very rare cases, like that of Otto, personal identity can be dependent on their environment in the *reidentification* sense, since, without his notebook, Otto presumably couldn’t identify himself, right now, with the person who went to MoMA yesterday, or got married forty years ago. Further, even though Otto is a special case, and most of us can reidentify ourselves without environmental cues, I believe that HEC might imply that personal identity in the *characterisation* sense might depend very heavily on our environment, and particularly other individuals, even in ordinary cases.

This kind of discussion is interesting for ethics because ‘person’ is generally a term of prudential and moral significance. I care about my future because I feel some special connection to a particular future person who I identify as myself, and not anyone else. Similarly, I care about my past actions because I feel a special connection to the person who undertook them. It is because of this ‘special connection’ that we hold others, and ourselves, to account for particular actions; it is because we deem those actions *theirs*. So there are some curious ideas that come out of this expansion of HEC into the area of personal identity, including the possibility that Otto is not responsible for a crime that he committed yesterday, assuming that he can't otherwise remember it, if we take away his notebook. Further, if our environment or other individuals have some responsibility for the identity of cognitively abled individuals in the *characterisation* sense, then it might be correct to allocate some blame for their actions to their surroundings, including other individuals, in a literal sense.

I will begin by discussing personal identity, taking inspiration from traditional literature, rather than HEC. I will conclude that psychological factors are crucial for our ability to identify ourselves as ourselves over time, and particularly, that it is our ability to place experiences within our own life narrative that gives us this sense of being the same person over time. Further, if these psychological factors are constituted by environmental factors, then the environment may, in part, be constitutive of our personal identity over time (or ‘diachronic identity’ herein).

2.1 Wherein does personal identity consist?

*2.1.1 The psychological criterion*

The assumption that underpins Lockean and neo-Lockean psychological accounts of diachronic identity is that:

“*X* at *t*1 is the same person as *Y* at *t*2 if and only if *X* is uniquely psychologically continuous with *Y*, where psychological continuity consists in overlapping chains of strong psychological connectedness, itself consisting in significant numbers of direct psychological connections like memories, intentions, beliefs/goals/desires, and similarity of character” (Shoemaker, 2012:§2.1)

So, following Locke, many advocates of the psychological criterion take *memory* to be critical for tracking a person through time: a person is “a thinking intelligent being, that has reason and reflection, *and can consider itself as itself*, the same thinking thing in different times and places; *which it does only by that consciousness which is inseparable from thinking*, and… essential to it” (Locke, 1836:225, emphasis added).[[5]](#footnote-5)

For Locke, personal identity is a ‘forensic’ notion – i.e. it has its home in courts of law. It is what allows us to praise, assign blame, or hold individuals to account for certain actions. To take Locke’s own example of the prince and the cobbler who swap brains, there seems to be something intuitively sound about saying that the prince continues to exist in the body of the cobbler, and the cobbler continues to exist in the body of the prince, because “whatever past actions [a person] cannot reconcile or appropriate to that present self by consciousness, [he/she] can be no more concerned in than if they had never been done” (Locke, 1836:234-5).

There are, however, some odd consequences of the Lockean psychological criterion, the most serious of which are that I cannot survive complex amnesia, and I am not responsible for any crime that I don’t remember committing. Not only might such claims be hard to swallow, but they also seem to involve a logical contradiction: the failure of transitivity of identity. Through Reid’s example of the brave officer (cited in Copenhaver, 2012), it is apparent that the Lockean psychological criterion is lacking: if I can't remember what I had for breakfast this morning, Locke would have to say that the person who ate my breakfast this morning wasn’t me. Yet although his is not a plausible theory on its own, the memory criterion does seem to be in some way related to the way in which I consider myself to have existed since breakfast this morning. Parfit’s augmentation of the Lockean criterion allows that, although there might be no “*direct memory connections*” (Parfit, 1987:205) between myself this evening and myself eating breakfast this morning, there may, nonetheless, be “*continuity of memory*” (Parfit, 1987:205), that is, there will be an overlapping chain of memories that tie together these otherwise psychologically distinct persons.

*2.1.2 The biological criterion*

Olson and others (DeGrazia, 1999; Olson, 1997a, 1997b; Carter, 1982) take issue with the psychological criterion because it seems to imply that *I* exist as long as (and *only* for as long as) my psychologically continuous states exist. But wasn’t I at one point a foetus? Or an infant incapable of forming memories? Mightn’t I one day find myself in a permanent vegetative state? On the psychological criterion, it would be incorrect to say that *‘I’* was, or could be, any of these things because *‘I’* cannot be psychologically continuous with any of them.

The alternative for the criterion for diachronic identity put forward by Olson *et al.* relates to biological make-up. This view takes the “structural integrity of an organism over time” (Wilson & Lenart, forthcoming:3) to be the deciding factor in continuity of personal identity. Olson (2003), for example, understands persons as essentially biological, or human, organisms; so, in order to be able to identify a person over time, we just need to establish the persistence conditions of human organisms. Easy enough. Personal identity, in this case, would be reducible to something like Lockean organismic identity, which is to do with the continuity of organisation of a living thing so that it is able to function as a single living thing over time, even through radical change (like the acorn becoming the oak).

On the biological criterion for personal identity, it seems that, after the brain-swap, the prince would continue to exist in the prince’s body and the cobbler would continue to exist in the cobbler’s body, even though there would be no psychological continuity whatsoever for either individual. But this certainly seems like a strange conclusion.

Things get worse still for Olson and his allies if we imagine an assailant and his victim switching brains (Wilson & Lenart, forthcoming:4). According to the biological approach, the assailant’s body should be punished for the crime, which means that the experience of punishment would be added to the victim’s stream of consciousness, which already also contains the traumatic experience of the attack. And yet, “[m]orally speaking, we ought to feel apprehensive about a view that would be willing to hold a person accountable even if its mind were unaware of the body’s previous activities and, in fact, cognizant of entirely different actions.” (Wilson & Lenart, forthcoming:4). I am not alone in believing that personal identity is strongly related to ethics, and supposing that any criterion for personal identity that doesn’t support our ethical intuitions must be problematic (see Shoemaker, 2012; Parfit, 1987). So, based on intuitions about this case, it seems that if we want to maintain the connection between persons and moral blame, as Locke wanted to, we have to move away from the biological criterion back towards the psychological.

*2.1.3 The non-reductionist view: souls and egos*

Our next option is to consider that there might be some substance (or self), which could be physical or immaterial, the diachronic identity of which might be *independent of its related psychology or biology*. Historically, this has been a popular position amongst philosophers, from Aristotle’s rational soul to the Cartesian Ego, and, despite being widely rejected within the more modern philosophical literature, it remains one of the most popular theories of identity *outside* philosophy. This view says that, when I say “I”, I refer to a persisting ego-substance (or soul), which is wholly present at every stage of my life, and which unifies every temporal slice of my life into the life of a single person.

However, if this independent self isn’t something that is identified with either biology or psychology, it isn’t clear how we could hope to track it over time and use it as a measure of diachronic identity. To use the words of David Shoemaker, “[w]e cannot track immaterial egos floating free from any particular psychological [or biological] properties, so on this view we would never be justified in claiming to have re-identified anyone” (2012:§2.5).

Another problem for the non-reductionist view is why my patterns of prudential concern are attached to a particular ego, rather than all of the relevant psychological features: what is it about this particular substance that warrants my concern? If it is by virtue of the fact that it carries various psychological connections, “then we might well wonder why we shouldn’t just care directly for those connections, rather than merely for the ‘package’ they come in” (Shoemaker, 2012:§2.5).

*2.1.4 Narrative identity*

Schechtman (1996) proposes a slight variation on the psychological view, which divides the issue into two parts: (1) the reidentification question: what are the conditions under which a particular person is properly reidentified at some other point in time? (2) The characterisation question: what are the conditions under which various psychological characteristics, experiences, and actions are properly attributable to a particular person (Shoemaker, 2012:§2.3)?

Narrative identity is therefore concerned with psychological unity, but it entails that the subject of this psychological unity is active in experience. In order for a subject of experience to be a genuine moral and prudential *agent*, it isn’t enough for her to simply have a set of experiences *happen* to her. Instead, “those experiences must be actively unified, must be gathered together into the life of one narrative ego by virtue of a story the subject tells that *weaves* them together, giving them a kind of coherence and intelligibility they wouldn’t otherwise have had” (Shoemaker, 2012:§2.3). It is by virtue of the fact that events happen within the context of a particular life-narrative that they have any meaning at all.

Narrative identity gives an adequate account of our prudential concerns, because persons are understood as temporally extended narrative egos, whose experiences form crucial parts of a particular story, rather than as temporal slices of egos that are, for some reason or other, concerned with other temporal slices of egos. It also gives a helpful account of responsibility, as what makes a particular action mine, rather than anyone else’s, is the fact that it fits coherently into my life-story, and is borne out of my beliefs, desires, and other experiences.

*2.1.5 Empirical evidence*

Theorising about which of the above possibilities is the essential criterion of personal identity is all well and good, but there is a wealth of empirical evidence to suggest that some sort of psychological criteria, including Schechtman’s characterisation criterion, is most in line with our practical experience of what it means to be the same person over time.

First, consider the effects of the *surgical division of the two hemispheres of the brain* (a ‘commissurotomy’), initially described by the Nobel Prize winning neuropsychologist, Roger Sperry (1964). His research led him to posit that “the surgery has left these people with two separate minds, that is, two separate spheres of consciousness. What is experienced in the right hemisphere seems to be entirely outside the realm of awareness of the left hemisphere” (299). Even though patients seemed to function relatively well in day-to-day life (since much of the activity that goes on in each hemisphere is similar enough), under certain circumstances the repercussions of the surgery become evident.

Sperry provides plenty of evidence that the left and right hemispheres, for these patients, really do perceive and comprehend entirely separately (1964:299-304) (although it is only under specialised conditions that a particular hemisphere can be isolated for testing, so the individual tends to be able to function in ordinary life as if they were unified). Sperry’s interesting conclusion, then, is that “there are presumably two separate systems perceiving, that is, conscious of the same thing, in the same way that two separate brains in two separate people look at and perceive the same event straight in front of them” (Sperry, 1964:303), although it is worth noting that this may be difficult to prove decisively, since one hemisphere is devoid of the capacity for expression. The idea that there are two separate cognitive systems (or persons) in play here is made clear by evidence that they sometimes conflict. One particular patient (“Case I”) reported that, while he was dressing, his right hand would try to pull his trousers up while his left would try to pull them down, or that the right hand might beckon to his wife to come over but, when she arrived, the left had would push her away (Sperry, 1964:304).

Given that there seem to be two distinct sets of desires, beliefs, and even experiences present within a single brain and body, it seems odd to suppose that they are unified as a single person, as the non-reductionist or biological accounts of personal identity would have to do. Neither approach, particularly the biological, could offer us very much in the way of explanation of these conflicting behaviours. A psychological criterion, however, could allow for this evidence, explaining it as an illustration of the existence of multiple persons within a single body.

Secondly, consider cases of *extreme memory loss over time*. Otto himself is characteristic of this interesting area of empirical research, which involves, for example, age-related dementia, and specifically Alzheimer’s disease. In these cases, the body remains (although, admittedly, dementia is caused by a loss of nerve cells in the brain) but, importantly, it is *the inability of sufferers to live a “cohesive mental life* [that] call[s] into question the relationship between one’s self at distinct times, such as the past and the present, or the present and the future” (Wilson & Lenart, forthcoming:8-9, emphasis added).

Studies have shown that carers of dementia patients feel as though their loved ones no longer exist, presumably because changes in personality and/or behaviour render the affected person altered or unrecognisable. Gillies and Johnston describe how carers are “acutely aware of the loss of their partner in terms of affection and togetherness and expressed a sense of emotional isolation from their relative”, as the patient reverts “to an almost childlike, asexual role as their increasing dependence insidiously [eats] away at their previous status and sexuality” (2004:439). One carer described her mother as “not my mother any more. That’s another pleasant wee woman who looks like her who I look after” (2004:440).[[6]](#footnote-6)

In these cases, the body remains, and yet the affected person seems no longer to be present in an important sense. In the experience of those around them, the dementia sufferer no longer displays many of those traits that were previously taken to be central to ‘who they are’; this is demonstrative of Schechtman’s *characterisation* sense of identity. Since Alzheimer’s is most closely associated with memory loss, there seems to be good reason to suppose that memory loss is linked to a loss of identity in this characterisation sense, as well as in the more obvious reidentification sense.

So studies of dementia can illuminate the path to favouring a psychological approach, and particularly highlight the importance of narrative memory for perceived maintenance of personhood. Further, Gillies and Johnston also unwittingly provide empirical evidence for the notion of extended personal identity, describing how the effects of dementia are mitigated in instances where the carer treats the sufferer as if their former self remains (for example, by giving them the authority to make certain decisions); the patient thereby maintains their character through reference to other people and objects.

2.2 Personal identity and HEC

So it seems that we are right to look at the mind for the part of us that constitutes our identity, in both the characterisation and reidentification senses. The psychological criterion alone is lacking because it is clear that persons are more than just their capacity to recall episodic memories; they are individuals with particular traits and the ability to arrange their experiences into a coherent life narrative. Further, the cognitively disabled actually retain some sense of identity in the collective remembering of others so, more than the regularly cognitively enabled, their personal identities become socially manifested properties.

Research shows that group-identity is positively related to well-being, suggesting that “group memberships provide the basis of a shared sense of social identification, furnishing individuals with the psychological and material resources to manage stressors more effectively” (Bevins, forthcoming:6; see also Haslam & Reicher, 2006; Thoits, 1983). Loss of identification with social groups is common in the aging individual as they may retire (losing their professional identity), their marital partner may die (losing their identity as a spouse), or they may generally become less mobile and able to participate in group interaction, particularly if they move into residential care. These changes constitute identity threats that are detrimental to the individual’s character and well-being, and they are all the more prevalent in people with dementia, whose functional and cognitive impairment make them even less able to sustain important group memberships. There are significant consequences for well-being in these cases, with as many as 50 per cent of dementia patients also suffering from depression (compared with around 20 per cent of otherwise healthy over-65s, and 40 per cent of over-85s) (Bevins, forthcoming:7).

So it seems as though there is a strong argument for the inclusion of group membership into our sense of diachronic identity. Since groups are necessarily made up of more than one individual, to some extent, *other individuals* seem to be included in our sense of identity. Further, by moving to a nursing home, individuals might take on more unwelcome group memberships, identifying themselves with other elderly or disabled people. Added to the unfamiliarity with their surroundings, which means that they are unable to cognitively offload many of their daily routine schedules “on items or places that *remind* them of what they have to do, how they ought to do it, and other kinds of pertinent information” (Wilson & Lenart, forthcoming:13), these people with memory-related disorders gradually lose their character and sense of identity much faster than they would if their group memberships and home surroundings were able to remain intact.

*2.2.1 Extended personal identity*

HEC brought about a new way of thinking about minds: it proposed that we often, under particular conditions, offload cognitive tasks onto our environment. Suppose that keeping track of our own diachronic identity is one of those tasks. In the same way that Otto remembered where MoMA is *because it is written in his notebook*, he can also recognise that the person who knows where MoMA is, is the same personas the person who, for example, ate scrambled eggs for breakfast or looks forward to meeting Inga for coffee next week, *because it is written in his notebook*.

And it isn’t just Otto to whom this notion of extended personal identity might apply.

His narrative memory, and therefore his sense of himself over time, is (almost) entirely constituted by his notebook because, without it, he wouldn’t be able to tell the story of his individual history. Yet, to a lesser extent, we all offload narrative memories onto our environment. For example, I remember what I was thinking about this paper when I was writing yesterday because I have it written down on this document. I am able to follow the train of thought that I had yesterday *because I offloaded it onto my laptop in a comprehensive way*. The person who was attached to the offloaded memory thereby continues to exist today because of this psychological continuity. As Lindemann argued, “[f]amiliar places and things, beloved objects, pets, cherished rituals, one’s own bed or favorite shirt, can and do help us to maintain our sense of self” (Lindemann, 2010:162-3).

And diachronic identity isn’t merely *extended*; it might also be *collective*. For example, social cognitive and social constructionist traditions in psychology have suggested that identity is constructed through conversations about the past, hence the well-established practice of ‘reminiscence work’ in elderly care, which is defined as “the discussion of past activities, events and experiences with another person or group of people” (Woods *et al.*, 2005:1). According to Mason *et al.* (2005), when cognitive decline prevents individuals from participating in particular social groups, a supportive group can play a crucial role in establishing social contact and reconstructing an identity based on social ties with others. The idea here is that “group reminiscence can provide a forum in which identity can be constructed and maintained as stories are shared between members,… leading to a shared sense of identity within a group context” (Bevins, forthcoming:12).

So externalism doesn’t just change how we think about minds, but also *persons*: selves are (partially) constituted by narrative memory, which, in most cases, is (partially) constituted by collective and extended processes. Thus, “*personhood* cannot be defined in solely individualistic terms” (Wilson & Lenart, forthcoming:17). If memory is distributed and offloaded, and therefore widely realised, it follows that diachronic identity is, too, widely realised.

*2.2.2 Ethical implications of extended personal identity*

If HEC entails a degree of extended personal identity, then there will be certain ethical implications, distinct from those discussed towards the end of chapter 1, which follow. The kinds of considerations that I am concerned with in this section relate to moral accountability and, specifically, retribution for particular criminal acts. Since retribution is usually delivered on the basis that the person receiving the punishment *is the same person* who committed the crime, there seem to be some interesting things that we could say if diachronic identity is widely constituted, even if it is only meaningfully so in a small range of cases involving extreme memory loss.

Motivation for retributive punishment is based on the idea that “offenders *deserve* punishment” (Dresser, 1990:420, emphasis added): the offender has committed a crime against society and has made it suffer, so society inflicts a proportionate amount of pain on the offender. The focus here is on the past, with the wrongdoing being what is important, rather than possible future gain. So, in order for “punishment to be fair, the person punished must be the same person who previously disobeyed the law” (Dresser, 1990:421). It seems, then, that the morally significant unit within the eyes of the law is the *whole succession of earlier and later selves*.

Consider the protagonist in the film, *Memento* (2000); he is the victim of a home invasion, during which he suffered a head injury and his wife was killed, and became unable to form new memories. The film is organised in segments, approximately ten minutes each in length, to represent the amount of time for which Leonard can remember anything new. The plot centres on his desire for and pursuit of revenge; he is looking for a “John G”, who he holds responsible for the crime, and leaves himself a set of clues to remind himself of any new evidence. However, as it transpires, the story that Leonard has told himself is fictional, and the viewer gradually learns that Leonard systematically misleads himself, leaving false clues, which direct him to kill innocent people along the way, including his best friend, Teddy. If we imagine that Leonard is arrested after murdering Teddy, it seems hard to imagine how we would justify punishing him, since there would be no person who has (1) the decision to kill Teddy, (2) the experience of killing Teddy, and (3) the experience of being arrested for Teddy’s murder, contained within his life narrative. On the conception of diachronic identity that I have advocated, there seems to be no good reason for retributive punishment.[[7]](#footnote-7)

However, the idea of extended personal identity might have some interesting implications for this kind of case. To return to (perhaps an altogether extreme version of) the Otto scenario: he too is unable to form or recall coherent memories so, without his notebook, he cannot place any of his actions, either past or present, within the context of an unfolding life narrative. Yet, as I have suggested, his personal identity could be understood as extending into his notebook, giving him access to the kind of life narrative that Leonard cannot have. Imagine now that Otto commits a crime; if there is a literal sense in which Otto *remembers* committing it *by virtue of having access to his notebook* (as HEC stipulates), then there does seem to be a person who has (1) the decision to commit the crime, (2) the experience of committing the crime, and (3) the experience of being arrested for the crime, contained within his life narrative, namely, the ‘Otto-plus-notebook’ person.

And so we come to a rather odd conclusion: that we would only be justified in punishing Otto, by putting him in prison, for example, if he could continue to have access to his notebook. If he didn’t, then the persons who committed the crime and received the punishment, respectively, wouldn’t be identical. HEC can therefore provide us with a distinction between Leonards and Ottos, perhaps even requiring us to provide the Leonards with unrestricted access to their former self (perhaps through journals or photo albums) if we want to be able to justify punishing them.[[8]](#footnote-8)

Another type of case for consideration requires us to revisit the idea of *socially extended cognition*. ‘Accomplice liability’ allows a court to find a person criminally liable for acts committed by a different person. In these cases, the court will examine the accomplice’s *intent* (whether or not they intended that the crime be committed), and the *scope of their liability* (the extent to which the accomplice provided aid, counsel, or encouragement towards the commission of the crime).

I propose that if we are reliant upon other persons and social groups in the sense that our personal identity is (partially) bound up with them, or that they are somehow (partially) responsible for our unfolding personality in the *characterisation* sense of personhood, then this might be a novel way to make sense of the intuition that underlies these points of law – namely, that individuals can be held to account for a crime that they did not, in practice, commit.

These kinds of considerations, motivated by HEC, also enable us to hold wider society to account for the actions of its members. There could perhaps be a literal sense in which various social groups are responsible for our character, and hence they are directly involved in the kinds of people that we become. For example, many of us have experiences of friends becoming entirely ‘different people’, perhaps after joining a new religious or political group. If other people (or society in general) constitute our personal identity to some degree, and a ‘person’ is what we deem to be a morally significant unit, then a person’s social groups must also be apportioned some of the responsibility for their actions, to whatever extent the group is responsible for shaping the individual.

**Conclusions**

In this paper I have tried to explore some of the implications of HEC, many of which I believe to have been neglected in the literature. I have demonstrated that the taxonomy that HEC provides is relevant to various areas of ethics, including neuroethics and moral accountability. It also provides us with a theoretical framework for our intuitions that wider society, or particular social groups, are responsible for the kinds of individuals that they produce. Further still, it prescribes a particular kind of treatment of cognitive artefacts, namely, that they should be treated as being on an ethical par with the brain.

I have also shown how the scope of HEC might be widened to include concerns about personal identity. Following on from the Lockean tradition, I have supposed that psychological factors are what matter, although I have noted Schechtman’s distinction between reidentification and characterisation as two separate senses of being the same person over time. Given that external artefacts sometimes constitute our memories, and that the people with whom we associate are partially responsible for shaping our character, it seems that diachronic identity “just ain’t in the head” (Putnam, 1973:704).

**Bibliography**

Works cited

Adams, F. & Aizawa, K. (2010), ‘Defending the Bounds of Cognition’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 67-80

Adams, F. & Aizawa, K. (2008), *The Bounds of Cognition*, Blackwell

Adams, F. & Aizawa, K. (2001), ‘The Bounds of Cognition’, *Philosophical Psychology*, 14(1), pp. 43-64

Allen-Hermanson, S. (2012), ‘Superdupersizing the Mind: Extended Cognition and the Persistence of Cognitive Bloat’, *Philosophical Studies*, 158(1), pp. 1-16

Anderson, J. (2000), *Leaning and Memory: An Integrated Approach*, (Second Edition), New York: John Wiley & Sons

Bevins, A (forthcoming), ‘Memory, Identity and Well-Being: Preserving Selfhood in Dementia’, available from <http://hdl.handle.net/10036/46973> [accessed 7/4/13]

Bobrow, S. & Bower, G. (1969), ‘Comprehension and the Recall of Sentences’, *Journal of Experimental Psychology*, 80, pp. 455-61

Burge, T. (1979), ‘Individualism and the Mental’, *Midwest Studies in Philosophy*, 4(1), pp.73-121

Carter, W. (1982), ‘Do Zygotes Become People?’, *Mind*, 91, pp. 77-95

Clark, A., (2010a) ‘Memento’s Revenge: The Extended Mind, Extended’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 43-66

Clark, A., (2010b) ‘Coupling, Constitution, and the Cognitive Kind: A Reply to Adams and Aizawa’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 81-100

Clark, A., (2009), ‘Spreading the Joy? Why the Machinery of Consciousness is (Probably) Still in the Head’, *Mind*, 118, pp. 963-993

Clark, A. & Chalmers, D. (2010), *The Extended Mind*, in Menary, R. (ed.), *The Extended Mind*, MIT Press: USA, pp. 27-42

Copenhaver, R., (2012) ‘Reid on Memory and Personal Identity’, *The Stanford Encyclopedia of Philosophy*, Zalta, E. (ed.), available from plato.stanford.edu/entries/reid-memory-identity [accessed 1/3/13]

DeGrazia, D. (1999), ‘Persons, Organisms, and the Definition of Death: A Philosophical Critique of the Higher-Brain Approach’, *Southern Journal of Philosophy*, 37, pp. 419-440

Dresser, R. (1990), ‘Personal Identity and Punishment’, *Boston Law Review*, 70, pp. 395-446

Ellison, R. (1995), *The Invisible Man*, 2nd ed., USA: Vintage International

Fleck, L. (1986), ‘Scientific Observation and Perception in General’, in Cohen, R. and Schnelle, T. (eds.), *Cognition and Fact: Materials on Ludwik Fleck*, Netherlands: Reidel Publishing Company, pp. 59–78

Gillies, B. & Johnston, G. (2004), ‘Identity Loss and Maintenance: Commonality of Experience in Cancer and Dementia’, *European Journal of Cancer Care*, 13, pp. 436-442

Harris, J. & Chatterjee, A. (2009), ‘Is it Acceptable for People to Take Methylphenidate to Enhance Performance?’, *BMJ*, available from <http://www.bmj.com/content/338/bmj.b1955> [accessed 17/4/13]

Haslam, S. & Reicher, S. (2006), ‘Stressing the Group: Social Identity and the Unfolding Dynamics of Responses to Stress’, *Journal of Applied Psychology*, 91(5), pp. 1037-1052

Hegel, G. (1977), *Phenomenology of Spirit*, translated by Miller, A., USA: Oxford University Press

Honneth, A. & Margalit, A. (2001), ‘Recognition’, *Proceedings of the Aristotelian Society, Supplementary Volumes*, 75, pp. 111-139

Hutchins, E. (2001), ‘Distributed Cognition’, *The International Encyclopedia of the Social and Behavioural Sciences*, pp. 2068-2072

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

Lakoff, G. & Johnson, M. (1999), *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*, New York: Basic Books

Latour, B. (1987), *Science in Action*, USA: Harvard University Press

Levy, N. (2007a), *Neuroethics: Challenges for the 21st Century*, New York: Cambridge University Press

Levy, N. (2007b), ‘Rethinking Neuroethics in the Light of the Extended Mind Thesis’, *The American Journal of Bioethics*, 7(9), pp. 3-11

Lindemann, H. (2010), ‘Holding One Another (Well, Wrongly, Clumsily), in a Time of Dementia’, *Cognitive Disability and its Challenge to Moral Philosophy*, Kittay, E. & Carlson, L. (eds.), Malden, MA: Wiley-Blackwell, pp. 161-168

Locke, J. (1836), *An Essay Concerning Human Understanding*, London: T. Tegg and Son, [online], available from books.google.co.uk [accessed 13/2/13]

Mason, E., Clare, L., & Pistrang, N. (2005), ‘Processes and Experiences of Mutual Support in Professionally-Led Support Groups for People with Early-Stage Dementia’, *Dementia*, 4(1), pp. 87-112

McNamara, D. S., & Healey, A. F. (1995), ‘A procedural explanation of the generation effect: The use of an operand retrieval strategy for multiplication and addition problems’, *Journal of Memory and Language*, 34, 399-416

Meyer, M. (2008), ‘Reflections on Comic Reconciliations: Ethics Memory, and Anxious Happy Endings’, *Journal of Aesthetics and Art Criticism*, 66(1), pp. 77-87

Mill, J. S. (2006), *On Liberty*, London: Adamant Media

Noë, A., (2007), ‘Magic Realism and the Limits of Intelligibility: What Makes Us Conscious?’, *Philosophical Perspectives*, 21, pp. 457-474

Olson, E. (2011), ‘The Extended Self’, *Minds and Machines*, 21(4), pp. 481-495

Olson, E. (2003), ‘An Argument for Animalism’, author manuscript available from eprints.whiterose.ac.uk/archive/00000734 [accessed 6/4/13]

Olson, E. (1997a), ‘Was I Ever a Fetus?’, *Philosophy and Phenomenological Research*, 58, pp. 95-109

Olson, E. (1997b), *The Human Animal: Personal Identity Without Psychology*, Oxford: Oxford University Press

Parfit, D. (1987), *Reasons and Persons,* 3rd edition, Oxford: Clarendon Press

Pöyhönen, S. (forthcoming), ‘Explanatory Power of Extended Cognition’, to appear in *Philosophical Psychology*, available from http://samulipoyhonen.files.wordpress.com [accessed 17/4/13]

Putnam, H. (1975), ‘The Meaning of ‘Meaning’, *Minnesota Studies in the Philosophy of Science*, 7, pp. 131-193

Redding, P. (2012), ‘Georg Wilhelm Friedrich Hegel’, *The Stanford Encyclopedia of Philosophy*, Zalta, E. (ed.), available from plato.stanford.edu/entries/hegel [accessed 11/4/13]

Roskies, A. (2002), ‘Neuroethics for the New Millennium’, *Neuron*, 35, pp. 21-23

Ross, D. & Ladyman, J. (2010), ‘The Alleged Coupling-Constitution Fallacy and the Mature Sciences’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 155-166

Rowlands, M. (1999), *The Body in Mind: Understanding Cognitive Processes*, Cambridge, UK: Cambridge University Press

Rupert, R. (2004), ‘Challenges to the Hypothesis of Extended Cognition,’ *Journal of Philosophy*, 101, pp. 389-428

Santayana, G. (1905), ‘Reason in Common Sense’, in *The Life of Reason*, New York: Schribners

Schechtman, M. (1996), *The Constitution of Selves*, Ithaca: Cornell University Press

Shoemaker, D. (2012), ‘Personal Identity and Ethics’, *The Stanford Encyclopedia of Philosophy*, Zalta, E. (ed.), available from plato.stanford.edu/entries/identity-ethics [accessed 28/3/13]

Sperry, Roger W. (1964), ‘Brain Bisection and Mechanisms of Consciousness’, in Eccles, J. (ed.), *Brain and Conscious Experience*, New York: Springer-Verlag, pp.298-313

Sprevak, M. (2009), ‘Extended Cognition and Functionalism’, *Journal of Philosophy*, 106(9), pp. 503-527

Suchman, L. (1987), *Plans and Situated Action*, Cambridge University Press

Sutton, J. (2010a), ‘Exograms and Interdisciplinarity: History, the Extended Mind, and the Civilizing Process’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 189-226

Sutton, J., Harris, C., Keil, P., Barnier, A. (2010b), ‘The Psychology of Memory, Extended Cognition, and Socially Distributed Remembering’, *Phenomenology and Cognitive Sciences*, 9(4), pp.521-590

Sweeting, H. & Gilhooly, M. (1997), ‘Dementia and the Phenomenon of Social Death’, *Sociology of Health and Illness*, 19, pp. 93-117

Thoits, P. (1983), ‘Multiple Identities and Psychological Well-Being: A Reformulation and Test of the Social Isolation Hypothesis’, *American Sociological Review*, 48(2), pp. 174-187.

Wheeler, M. (2010), ‘In Defense of Extended Functionalism’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 245-270

Whetstone, T. & Cross, M. (1998), ‘Control of Conscious Contents in Directed Forgetting and Thought Suppression’, *PSYCHE*, 4(16), pp. 4-16

Wilson, R. & Lenart, B. (forthcoming), ‘Extended Mind and Identity’, draft 3.0, to appear in *Handbook of Neuroethics*, Clausen, J. & Levy, N. (eds.), Springer, author manuscript available from www.artsrn.ualberta.ca/raw/ [accessed 4/4/13]

Wilson, R. & Foglia, L. (2011), ‘Embodied Cognition’, *The Stanford Encyclopedia of Philosophy*, Zalta, E. (ed.), available from plato.stanford.edu/entries/embodied-cognition [accessed 11/4/13]

Wilson, R. (2004), *Boundaries of the Mind: The Individual in the Fragile Sciences*, Cambridge, UK: Cambridge University Press

Woods, B., Spector, A., Jones, C., Orrell, M., Davies, S. (2005), ‘Reminiscence Therapy for Dementia’, *Cochrane Database of Systematic Reviews*, 2, available from <http://onlinelibrary.wiley.com> [accessed 10/4/13]

Other works consulted

Baker, L. (2009), ‘Persons and the Extended-Mind Thesis’, *Zygon: Journal of Religion and Science*, 44(3), pp. 642-658

Clark, A. (2008), *Supersizing the Mind*, Oxford University Press

Dewhurst, J. (2012), ‘The Ethics of Cognitive Extension’, *Aporia*, 9, pp. 22-28

Lau, J. and Deutsch, M. (2010), ‘Externalism about Mental Content’, *The Stanford Encyclopaedia of Philosophy*, Zalta, E. (ed.), available from plato.stanford.edu [accessed 27/2/13]

Levin, J. (2010), ‘Functionalism’, *The Stanford Encyclopedia of Philosophy*, Zalta, E. (ed.), available from plato.stanford.edu/entries/functionalism [accessed 16/4/13]

Menary, R. (2010a), ‘Introduction: The Extended Mind in Focus’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 1-26

Menary, R. (2010b), ‘Cognitive Integration and the Extended Mind’, in Menary, R. (ed.), *The Extended Mind*, Cambridge, USA: MIT Press, pp. 227-244

Puccetti, R. (1973), ‘Brain Bisection and Personal Identity’, *British Journal for Philosophy of Science*, 24, pp. 339-355

Shoemaker, D. (2000), ‘Reductionist Contractualism: Moral Motivation and Expanding the Self’, *Canadian Journal of Philosophy*, 30(3), pp. 343-370

Shoemaker, D. (1999), ‘Selves and Moral Units’, *Pacific Philosophical Quarterly*, 80(4), pp. 391-419

Tajfel, H. & Turner, J. (1986), ‘The Social Identity Theory of Intergroup Behaviour’, in Worchel, S. & Austin, W. (eds.), *Psychology of Intergroup Relations* (pp. 7-24), Chicago: Nelson-Hall, pp. 7-24

Turner, J., Oakes, P., Haslam, S., & McGarty, C. (1994), ‘Self and Collective: Cognition and Social Context’, *Personality and Social Psychology Bulletin*, 20, pp. 454-463

Audiovisual media

*Eternal Sunshine of the Spotless Mind*, (2004), [Film], directed by Michael Gondry, USA: Universal Studios

*Memento*, (2000), [Film], directed by Christopher Nolan, USA: Sony Pictures

1. This is an allusion to Clark’s Parity Principle. See section 1.3 for details. [↑](#footnote-ref-1)
2. I allude here to A&A (2001), where they claim the mark of the cognitive to be “particular kinds of processes involving non-derived representations” (52-3). [↑](#footnote-ref-2)
3. This might seem like rather a significant concession to make: why defend HEC if HEMC can do all of the same work in the neuroethical arena? I remind the reader that the aim of my project is not to evaluate the relative merit of HEC and HEMC, but to explore the implications of HEC, which must include neuroethical considerations. My ‘concession’ here to HEMC is intended to be more like a safety net, ensuring that these issues aren’t overlooked by the staunch contra-HEC theorist. [↑](#footnote-ref-3)
4. For any reader who hasn’t seen the film, the basic premise is that it is possible to erase *particular individuals* from one’s memory. [↑](#footnote-ref-4)
5. Note that “consciousness” here would be better not taken as being of the phenomenal kind, because this, as I explained in section 1.1, is a contentions issue within HEC. I interpret Locke here to be referring to the kind of psychological continuity referred to in the Shoemaker quote above. [↑](#footnote-ref-5)
6. Sweeting and Gilhooly have described this phenomenon as a “social death” (1997:93). [↑](#footnote-ref-6)
7. However, punishment is sometimes motivated by other factors, such as deterrence, incapacitation, and rehabilitation, any of which might be appropriate for Leonard. [↑](#footnote-ref-7)
8. This might be taken to be a *reductio ad absurdum* of HEC; I expect that this will depend on the reader’s intuition about HEC, and which side of the Rowlands deadlock he/she finds himself on. As I explained in the introduction, I will have to leave this issue fairly open-ended. [↑](#footnote-ref-8)