Extended Memory, the Extended Mind, and the Nature of Technology-Mediated Memory Enhancement

Paul R. Smart
School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK.

Introduction

Human memory is a key focus of scientific and theoretical attention within a range of disciplines, and it is also an important target of efforts that seek to improve aspects of human cognitive function. The phenomenon of memory is also something that has been seen as a key test for theories of distributed, situated and extended cognition, not least because memory seems to take us out of the current situation and put us in psychological contact with a set of previously experienced state-of-affairs. As Sutton (2009) comments, "The challenge set by the nature of human memory to theories of situated cognition...is to see how social or material resources outside the brain could possibly be an integral or constitutive feature of memory states or processes, when the events or episodes remembered are long gone" (pg. 229). This is deemed to pose a challenge to theories of cognition that emphasize the importance of extra-neural structures and resources in scaffolding, sculpting, and (in the case of extended cognition) realizing cognitive processes (Clark, 2008; Robbins & Ayded, 2009).

In addition to the challenge posed to extended cognition accounts, memory also raises interesting issues when it comes to the potential impact of certain new techniques and technologies, all of which seem poised to exert some effect on our mnemonic functioning. For example, the increasing availability of life-logging technologies (O'Hara et al., 2009), coupled with the increasing use of the Web as a storage medium for personal data, raises important questions about the potential of the Web to 'enhance' our mnemonic capabilities. A critical question here concerns the extent to which we can treat Web-based resources as counting as part of our memory system. For example, if we are enabled to have more-or-less immediate, reliable, and easy access to bodies of information on the Web, does that information count as part of *our* own body of knowledge and beliefs about the world – in the same way, perhaps, as the content of our biologically-based (semantic) memories? If it does, then it seems we are only a few technological steps away from an era in which the limits of our personal knowledge are defined by the extent of the Web's reach.

The aims of this particular paper are threefold. The first aim is to review the literature relating to cognitive extension and the extended mind and to illustrate how these ideas are relevant to the case of memory. A second aim is to consider the value of an extended cognition account in thinking about memory phenomena. The focus here is on the role played by external (non-biological) physical and social resources in shaping our mnemonic capabilities. A third and final aim for the paper is to consider a variety of issues related to the design of memory technologies. Important areas of discussion here include the extent to which memory technologies should aim to support the *accurate* recall of previously experienced information, as well as the role of biology in guiding the design of memory technologies.

Movies, Memories and the Extended Mind

Memento

Human memory features as a key focus of attention in a broad range of academic disciplines. Biologists, psychologists, philosophers, sociologists, media theorists, computer scientists, and others: all have contributed in one form or another to our understanding (or misunderstanding) of memory. However, it is not just the scientific and philosophical community that have expressed an interest in memory. Memory has also been a focus of attention in the literary and cinematic community. In the case of cinema, films such as *Blade Runner*, *The Eternal Sunshine of the Spotless Mind* and *Total Recall* are all films in which memory features as a key thematic element. However, perhaps the most compelling cinematic treatment of human memory in recent years, and the one that is most relevant to the current paper, is Christopher Nolan's *Memento*. There are a number of ways in which *Memento* is relevant to the present discussion, but first it will help to provide an overview of what the movie is actually about.

Memento tells the story of Leonard (Lennie) Shelby, a former insurance investigator, who suffers from a form of anterograde amnesia. Lennie's condition results from a head trauma that was sustained while attempting to defend his wife from an attack some time before the events of the movie take place. We are led to believe that during the attack, Lennie's wife was raped and killed, and that Lennie, driven by a desire for vengeance, is intent on finding and killing the person responsible for his wife's death. Here, however, Lennie faces a problem. The nature of his memory impairment makes it difficult for him to form new memories. His long-term memories leading up to the attack are unaffected, but since the attack on his wife Lennie is unable to form any new long-term memories. Lennie's life is thus lived as a series of short, disjointed episodes, each supported only by the capacity of his short-term memory[1]. The only way in which Lennie can make any sense of his ongoing experiences, is by constantly referring to an array of self-generated resources, such as hand-written notes, annotated photos, and body tattoos. These provide some sort of historical context for the situations Lennie finds himself in, and they also serve to guide his thought and actions in (hopefully) goal-appropriate ways. Although the resources Lennie relies on seem to serve as poor substitutes for biological memory, Lennie is undeterred by the seemingly ersatz nature of his 'external memories'. In fact, he believes his system of storing information in external resources is actually an improvement on biological memory: "Memory's unreliable...Memory's not perfect," says Lennie, "It's not even that good. Ask the police; eyewitness testimony is unreliable...Memory can change the shape of a room or the color of a car. It's an interpretation, not a record. Memories can be changed or distorted, and they're irrelevant if you have the facts." By systematically exploiting the resources he creates, Lennie believes he will eventually be able to achieve his aim; he will eventually be able to piece together the mystery of his wife's death, track down his wife's killer, and exact his revenge.

Memento is a movie about many things besides memory – trust, grief, identity, knowledge, moral responsibility, agency, etc. – and some of these topics surface (albeit superficially) throughout the

[1] The nature of Lennie's fractured consciousness is reflected in the narrative structure of the film, which features two inter-leaved chronologies, one running forward, the other backward. The result, for the viewer, at least on a first viewing, is somewhat bewildering. The narrative structure of the movie enables the viewer to identify with Lennie both subjectively and epistemologically. Like Lennie, the viewer is left trying to make sense of each scene with a limited amount of background information to support their interpretation.

remainder of this paper. One thing, however, that is of key significance to the present discussion is the method used by Lennie to cope with his memory impairment; i.e. his systematic creation, maintenance and exploitation of a rich array of information-bearing artefacts (i.e. the aforementioned notes, photos and tattoos). These resources arguably function in a way that is reminiscent of our biological memories: they serve to guide and structure Lennie's thoughts and actions in a way that is similar to that served by biological memory. To the extent this is true, we might be inclined to say that some of Lennie's mnemonic capabilities are no longer solely supported by the machinations of Lennie's brain. Instead, we might be inclined to say that some of the machinery underlying Lennie's memory now extends beyond the biological borders of the brain to include elements of the external, (self-created) informational environment that Lennie inhabits. In other words, we could say that parts of Lennie's mind no longer *solely* depend on the whirrings and grindings of Lennie's damaged brain. Lennie's mind is now an *extended mind*. It is a mind that depends on an environmentally extended nexus of representational and computational elements (see Smart et al., 2010).

Such, at least, is the view suggested by a recently influential approach to mind and cognition, first proposed by Clark and Chalmer's (1998). The approach goes via a variety of names, such as locational externalism (Wilson, 2000, 2004), active externalism (Clark & Chalmers, 1998), vehicle externalism (Hurley, 1998; Rowlands, 2006), environmentalism (Rowlands, 1999), cognitive extension (Clark, 2008) and the extended mind (Clark & Chalmers, 1998). What all of these locutions have in common is a commitment to the idea that aspects of the external, extra-neural environment can, at certain times, play a constitutive role in the material realization of the mind (or aspects thereof). If such a view is true, then the implications for our understanding of memory and the role played by memory technologies in shaping, scaffolding, and perhaps even supplanting our biological memories are indeed profound. Before we move onto such issues, however, it is worth fleshing out the notions of cognitive extension and the extended mind in a little more detail.

Cognitive Extension

Consider the webs of the garden cross spider, *Araneus diadematus*. The webs, like that of most spider's webs, are architecturally complex, composed as they are of multiple types of silk thread, each laid down in a specific sequence and geometric pattern. Specific types of thread need to be produced at just the right time, and the overall design of the web has to be sensitive to a number of factors including the size of the prey to be caught and the shape of the local environment (the shape made available by local branches or other supporting structures). Presumably, in order to build its web, the spider needs some ability to remember the specific steps involved in constructing a web, and while constructing the web, it needs to know exactly where it is in the overall process in order to know what to do next. The former ability, we might say, is the spider's long-term memory (the spider's memory of the steps required to build a web), while the latter ability is the spider's working memory (the spider's memory of task-relevant information that is used to guide decision-making in goal-appropriate ways). So, when it comes to web construction spiders seems to be pretty sophisticated critters, and this view is reinforced when we think of a range of other behaviours that spiders are capable of, from optimal route selection in the case of the araneophagic spiders (Barrett, 2011, chapter 4; Tarsitano & Jackson, 1997) through to deceptive signalling in the case of jumping spiders (Wilcox & Jackson, 2002). The problem is that the spider's capacity for planning, decision-making, memory, judgment and problem-solving seem to be out of kilter with our intuitions about the complexity of the mechanistic substrate required to support such abilities. As invertebrates, spider's do not have a real brain; instead, their central nervous system is composed of a number of ganglia (collections of neural tissue), of which the most prominent are the supraesophagal and subesophagal ganglia. These ganglia serve to implement and coordinate the majority of the spider's sensorimotor functions. The surprising thing is that the total number of neurons in the spider's central nervous system is relatively small, at least when we compare it to the 100 billion neurons of the average human brain. The wandering spider, *Cupiennius*, for example, has 100,000 neurons, while the orb web spider, *Argiope*, has a nervous system composed of a mere 30,000 neurons (Foelix, 1996). These numbers do not seem particularly large given the kind of behaviours spiders are capable of. And they certainly do not seem sufficient for us to attribute cognitive capabilities like planning, memory, decision-making and so on. The machinery of the mind, in the case of the spider, seems far too small and insufficiently complex to warrant the spider's treatment as a cognitive agent. And yet the spider clearly functions very well in the world, despite its rather meagre set of neurocomputational resources. Despite having a nervous system that features fewer neurons than are encountered in some artificial neural networks, spiders seem very adept at finding intelligent solutions to challenging problems. How exactly do they do it?

The answer, at least in the case of web construction, seems to lie not so much in the details of the spider's nervous system, as in the nature of the spider's interaction with the external (extra-neural) environment. In particular, a detailed ethological examination of web spinning behaviour suggests that spiders are sensitive to a range of bodily contingencies involving the relative positioning of their legs on certain types of silk thread (Krink & Vollrath, 1997, 1998, 1999). As the web develops, the positioning of the spider's legs becomes a reliable cue as to what type of action needs to be executed next, as well as what type of silk needs to be produced. In essence, the web serves as "its own best model" (see Brooks, 1991) of what needs to be accomplished, and the spider need only be responsive to local information concerning the structural organization of threads in the immediate vicinity of its body. At each stage of the web construction process, each of the spider's legs need only perform a local (spatial) search for the nearest thread, and, once located, the relative positioning of the legs (as well as the type of thread they are in contact with) 'represents' the web's structural status. In response to this rich body of local information, the spider need only implement locally-effective rules concerning which action to perform. And it turns out that aspects of spider web weaving behaviour can be modelled using a relatively simple (and minimal) set of rules (Krink & Vollrath, 1997, 1999). Importantly, each rule exploits facts about the spider's bodily design, and its outputs specify actions that are geared to structuring the problem space in ways that guide, constrain and structure subsequent behaviour. The spider, it seems, distributes the computational burden associated with web spinning behaviour across a complex system that comprises its brain, body and aspects of the (self-structured) external environment[2].

The moral of this story, then, is that it is easy to be misled into thinking that intelligent action is always the sole product of neural mechanisms – that the point source of intelligent behaviour is always something that must reside in the 'head' of an agent. For what the case of web construction teaches us is that agents may often co-opt a variety of far flung forces and factors into a problem-solving routine, and not all of these forces and factors need be biological in nature. When seen in a

[2] Its body is (perhaps non-accidentally) designed so as to best exploit this state of affairs – you can represent quite a lot of information when your representational repertoire is sensitive to the spatial dynamics of a system comprising eight articulated appendages!

certain light, the external environment emerges as more than just a space for sensory inputs and motor outputs; it is also poised to play an important (explanatorily-potent) role in the mechanisms by which that behaviour is realized. Intelligent behaviour, we might say, is at least sometimes realized by processing loops that *extend* beyond the neural realm and productively incorporate a variety of extra-neural resources. Some forms of behavioural intelligence are *environmentally-extended* with regard to their mechanistic realization.

Accounts like this, which emphasize the contribution of bio-external factors in the mediation of intelligent behaviour, are relatively common in the cognitive scientific literature, and they are typically cases of what Wheeler and Clark (Wheeler, 2005; Wheeler & Clark, 1999) have referred to as 'non-trivial causal spread'. According to Wheeler and Clark, we encounter cases of non-trivial causal spread whenever we confront a phenomenon that has the initial appearance of being the product of a well-demarcated system, but which, on closer inspection, turns out to involve the exploitation of a variety of other forces and factors. And whenever we have a case of non-trivial causal spread, we also have a case of explanatory spread; i.e. a relative expansion of our explanatory frameworks to account for the phenomenon in question. Such spread seeks to give explanatory weight to factors that were initially supposed to be causally-irrelevant with respect to some target phenomenon, and in cases where the target phenomenon is a cognitive process then it seems to make sense to see the causally-active physical vehicles of the process as extending beyond the inner, neural realm. Inasmuch as we equate the boundaries of a cognitive system with the physical limits of the mechanisms that comprise that system's cognitive processing routines, then cognition is, at least sometimes, not bounded by the traditional borders of skin and skull; it emerges as something that is perfectly able to extend beyond the head and seep into the world.

Cases of non-trivial causal spread are evident throughout much of the distributed, situated, embodied and extended cognition literatures (Clark, 1997; Haugeland, 1998; Wilson, 2002), and they are also increasingly evident in the work of roboticists and ethologists (Barrett, 2011; Dawson et al., 2010; Pfeifer & Bongard, 2007; Pfeifer & Scheier, 1998). As a further example of non-trivial casual spread, Clark (1997) cites the example of the bluefin tuna, which is able to accomplish feats of swimming efficiency that seem to surpass that deemed possible based solely on a consideration of the tuna's biomechanical profile (the parallels with the case of web construction, in terms of a mismatch between the apparent biological make-up of the organism and the behavioural performances it is capable of, should be obvious). At first blush, it might appear necessary to assume that the tuna's swimming capabilities are necessarily grounded in the details of the tuna's physiology (its streamlined shape and powerful muscles), and that there must be some aspect of the tuna's physiology that has been overlooked. This, however, is not the case. Tuna, as it turns out, use their biological swimming apparatus to create and then exploit vortices in the water, and it is this that accounts for their otherwise mysterious natatorial capabilities (Triantafyllou & Triantafyllou, 1995). As with the case of spider web construction, what we see here is a case of brain, body and world working in concert to achieve performances that would otherwise be difficult or impossible to achieve if left entirely to biological elements. As Clark notes, the real 'swimming machine', in the case of the bluefin tuna, is not simply the biologically-bounded tuna but the tuna in its proper environmental context – the tuna plus the aqueous medium that it exploits to its own ends. Likewise, we can say that in the case of web construction the real 'spinning machine' is not simply the spider but the spider plus the environmental structures that it creates and subsequently exploits in the context of its web spinning activity.

All of this, I suggest, encourages us to question whether the machinery of cognition is solely biological (neural) in nature. It is common for us to see mental states and processes as having their material roots in whatever it is the brain does, and, in fact, this notion is now so much a part of our intellectual heritage that it is difficult for us to see things any other way. Yet just because the brain plays an undoubtedly important role in cognition, this does not mean that we can fully understand cognition by focusing all of our explanatory attention on the brain alone. Perhaps, as in the case of swimming and spinning, our understanding of cognition should be grounded in an appreciation of the way in which the biological brain works in concert with a variety of other, non-biological resources. Inasmuch as this is true, then we may need to broaden our explanatory focus beyond the neural realm – we may need to see the 'cognitive machine' (Smart, 2010) as something more than that which resides solely within the head.

The Extended Mind Hypothesis

The previous section highlighted the way in which certain types of intelligent behaviour and cognitive processing seem to include (as wholes do their proper parts) mechanisms that extend beyond the traditional biological borders of skin and skull. The specific claim is that, under at least some conditions, we are warranted in seeing cognition as, quite literally, extending into the extra-organismic environment. The argument as currently presented, however, might be seen as applying to a narrow subset of mental states and processes, at least relative to those that we typically associate with a human mind. In accounting for much of the behaviour of both ourselves and others we typically make reference to a set of common-sense, mentalistic terms (such as belief, desire, hope, fear, and so on), and these are seen as playing a genuine explanatory role in psychologically-interesting patterns of behaviour. Thus, my action to retrieve a cookie from the pantry is explained in terms of my 'desire' to eat a cookie and my 'belief' that cookies are to be found in pantries. It is this kind of intentional characterization (the ascription of intentional mental states) that helps us make sense of (to understand) patterns of human behaviour – it enables us to gain a predictively and explanatorily potent toehold on patterns of behaviour that would otherwise be psychologically unintelligible to us. So the question that arises in the case of cognitively-extended systems is whether the notion of cognitive extension gains any purchase in the more ethereal domain of folk-psychological discourse (the strategy of explaining human behaviour with respect to mental states, such as belief and desire). Can the notion of cognitive extension, as currently presented, be extended to account for the mental states that are posited as causally-relevant to the psychological understanding of our everyday patterns of behaviour? Can we, in other words, extend the case of an environmentally-extended cognitive system to the more general case of an environmentally-extended mind?

It is here (perhaps not surprisingly) that the philosophical waters begin to run deep. Perhaps the most lucid and influential account of why we should take notions such as extended belief states seriously is provided by Clark and Chalmers (1998) in their classic paper, 'The Extended Mind'. Clark and Chalmers (1998) ask us to imagine two individuals: Inga and Otto, both of whom are situated in New York City. Inga is a normal human agent with all the usual cognitive competences, but Otto suffers from a mild form of dementia and is thus impaired when it comes to certain acts of information storage and recall. To attenuate the impact of his impairment on his daily behaviour, Otto relies on a conventional notebook which he uses to store important pieces of information. Otto is so reliant on the notebook and so accustomed to using it that he carries the notebook with him wherever he goes and accesses the notebook fluently and automatically whenever he needs to do so. Having thus set the stage, Clark and Chalmers (1998) ask us to imagine a case where both Otto and Inga wish to visit the Museum of Modern Art (MOMA) to see a particular exhibition. Inga thinks for a moment, recalls that the museum is on 53^{rd} street and then walks to the museum. It is clear that in making this episode of behaviour intelligible (or psychologically transparent) to us Inga must have *desired* to enter the museum, and it is clear that she walked to 53^{rd} street because she *believed* that that was where the museum was located. Obviously, Inga did not believe that the museum was on 53^{rd} street in an occurrent sense (i.e. she has not spent her entire life consciously thinking about the museum's location); rather, she entertained the belief in a dispositional sense. Inga's belief, like perhaps many of her beliefs, was sitting in memory, waiting to be accessed as and when needed.

Now consider the case of Otto. Otto hears about the exhibition, decides to visit the museum, and then consults his notebook to retrieve the museum's location. The notebook says the museum is on 53^{rd} street, and so that is where Otto goes. Now, in accounting for Otto's actions we conclude, pretty much as we did for Inga, that Otto *desired* to go to the museum and that he walked to 53^{rd} street because that is where he *believed* the museum was located. Obviously, Otto did not believe that the museum was on 53^{rd} street in an occurrent sense (Otto has not spent much of his life constantly looking at the particular page in his notebook containing museum-related facts); rather, he entertained the belief in a dispositional sense. Otto's belief, like perhaps many of his beliefs, was sitting in the notebook, waiting to be accessed as and when needed.

Clark and Chalmers (1998) thus argue that the case of Otto establishes the case for a form of externalism about Otto's states of dispositional believing. The notebook, they argue, plays a role that is functionally akin to the role played by Inga's onboard bio-memory. If this is indeed the case, then it seems to make sense to see the notebook as part of the material supervenience base for some of Otto's mental states, specifically his states of dispositional belief (such as those involving museum locations). The main point of the argument is to establish a (potential) role for external artefacts in constituting the physical machinery of at least some of our mental states and processes. If, as Clark and Chalmers (1998) argue, the functional contribution of an external device is the same as that provided by some inner resource, then it seems unreasonable to restrict the material mechanisms of the mind to the inner, neural realm. It seems possible, at least in principle, for the human mind to occasionally extend beyond the head and into the external world.

Such claims are, understandably, disconcerting, and it is important that we understand the precise nature of the claim that is being made. One immediate cause for concern relates to the notion of functional equivalence between the inner (e.g. bio-memory) and outer (e.g. notebook) contributions. If we allow any form of externally-derived influence to count as part of the mechanistic substrate of the mind, then doesn't this cast the mechanistic net too widely? Don't we end up confronting cases that are so blatantly counter-intuitive that they undermine the very notion of the mind as a proper focus of scientific and philosophical enquiry? Consider, for example, the case where two people have a conversation on the bus. Does this mean that their respective minds have merged into one integrated whole? And what about cases where we have some very loose coupling with an external information source, say the kind of access we have to information in a conventional textbook? Clearly, not all of the technologies or external resources that we encounter are apt to engage in the kind of bio-technological hybridization envisioned by the extended mind hypothesis. As Clark (1997) argues:

> *"There would be little value in an analysis that credited me with knowing all the facts in the Encyclopaedia Britannica just because I paid the monthly installments and found space for it my garage" (pg. 217).*

Similarly, it would be foolish to equate my personal body of knowledge and beliefs as co-extensive with the informational contents of the internet simply because I have an internet-enabled mobile phone. What, then, are the conditions under which we count a set of external resources as constituting part of an environmentally-extended mind? In answering this question, Clark and Chalmers (1998) embrace a particular set of criteria, ones that appeal to the accessibility, portability, reliability and trustworthiness of the external resource. The criteria are that:

1. "…the resource must be available and typically invoked" (Clark, 2010). **[Availability Criterion]**
2. "…any information…retrieved from [the non-biological resource must] be more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory" (Clark, 2010). **[Trust Criterion]**
3. "…information contained in the resource should be easily accessible as and when required". (Clark, 2010) **[Accessibility Criterion]**

Clearly, such criteria serve to guide and constrain our intuitions about the kind of bio-artifactual and bio-technological couplings that are relevant to the formation of an extended mind. And they do so precisely because they delimit the range of situations under which we recognize the capabilities engendered by an external resource as being (most plausibly) that of a specific individual (or agent). In other words, what is important about the various criteria Clark and Chalmers (1998) propose is that they ensure that the capacities of an environmentally-extended, bio-technologically hybrid system are most plausibly seen by external observers (and perhaps by the agents themselves) as the capacities and features of a particular agent. As Wilson and Clark (2009) suggest:

> *"We properly expect our individual agents to be mobile, more or less reliable, bundles of stored knowledge and computational, emotional and inferential capacities. So we need to be persuaded that the new capacities enabled by the addition of the notebook are likewise sufficiently robust and enduring as to contribute to the persisting cognitive profile we identify as Otto the agent. The bulk of Clark and Chalmers' (1998) work was an attempt to isolate and defend a specific account of the conditions under which we would be justified in identifying such an extended mind." (pg. 67).*

What we seem to confront, then, is a set of what might generally be referred to as *coupling conditions*, conditions that determine when we are and when we are not justified in identifying cases of cognitive extension that apply to the realm of folk-psychological theorizing. In all cases of cognitive extension, what seems to be important is a particular pattern of temporally fine-tuned information flow and influence within a networked ensemble of diverse resources. This network constitutes the mechanistic substrate of an extended *cognitive* system whenever the objective of that system, or the task in which it is engaged, is recognizably cognitive in nature. However, this networked ensemble need not be permanent in nature. It can be a one-off organization that is assembled for the purposes of a specific cognitive task, or it can be a temporary but repeatable organization that is assembled to deal with an intermittent or periodically-occurring task (see Wilson

& Clark, 2009). When the organization is more permanent, we approach the kind of conditions under which we count the external resource as constituting part of the material supervenience base associated with an agent's daily patterns of psychologically-interesting behaviour. These are precisely the kind of conditions under which we are justified in seeing the emergence of an environmentally-extended mind.

The thought experiment invoked by Clark and Chalmers to support their notion of an extended mind raises important questions about the possibility of environmentally-extended memory systems. In particular, the way in which Otto uses his notebook seems to justify its inclusion as part of his cognitive apparatus – the notebook is essentially serving as a means by which some of Otto's memory functions can be implemented. The notion of cognitive extension also appears applicable to the case of Lennie in the movie, *Memento*. Lennie relies on a broader array of resources than Otto; however, his behaviour, like Otto's, is guided by the content of external resources in pretty much the same way as information retrieved from bio-memory[3].

Obviously, Otto and Lennie are fictional cases, so one question that begins to loom large hereabouts is the extent to which the notion of the extended mind is applicable to real-world cases of memory function? Does the extended mind thesis gain any traction when it comes to a consideration of human memory as encountered in everyday situations? Furthermore, what are the implications of the extended mind thesis for the design of memory technologies? The answers to these two questions are ones that will occupy us for the remainder of the paper.

Memory in the Real World

One way to press the case for an extended mind approach to memory is to consider the nature of memory as we typically encounter it in the real world. Of course, it is quite possible that all forms of memory, as currently constituted, are not extended and do not rely on external elements for their mechanistic realization. Even if this were so, it would not rule out the possibility of an extended approach to future forms of memory: ones in which technological artefacts are used to work in concert with the biological brain in order support (or supplant) bio-memory capabilities. Nevertheless, I think there are grounds for adopting distributed, situated or extended perspectives when it comes to our understanding of everyday mnemonic functioning as it is *currently* constituted. In the sections that follow, I present a number of cases where a consideration of the facts of material embodiment and environmental embedding can contribute to our understanding of memory phenomena.

Exploiting Environmental Structure

Perhaps the most obvious way in which the environment is employed in the service of memory is evident in those cases where we structure the environment in ways that meliorate memory

[3] An interesting issue here concerns the extent to which Lennie fulfils the coupling conditions proposed by Clark for extended cognitive systems. These are the conditions of availability, trust and accessibility. Certainly, Lennie fulfils the criteria of availability: he carries his notebook, photos and (obviously) his tattoos with him wherever he goes, and he relies on them in most situations. He also seems to fulfil the trust condition. In fact, for reasons that will become clear later in the paper, it is his trust in the information retrieved from external resources that is the source of his undoing. Finally, regarding the issue of accessibility, Lennie accesses his resources more or less automatically, and most of his resources are easily accessible to him. Without question, the resources are not as accessible as bio-memories, but it is not clear that this undermines their status as part of Lennie's 'memory machine'.

processes. Examples include the deliberate marking or tagging of items to reduce memory load (see Dennett, 1996, chapter 5), as well as the creation of cues (photos, notes, and so on) to assist us with retrieval operations. These activities are typically explained in terms of the offloading of memory onto the physical environment. As Dennett (1996) writes:

> *"Putting deliberate marks on the environment...is an excellent way of reducing the cognitive load on your perception and memory."* (pg. 135)

One particularly compelling example of how environmental structure is used to reduce cognitive load is provided by Beach's (1988) analysis of how expert bartenders remember a lengthy drinks order. The study (discussed in Clark (2008)) emphasizes how bartenders rely on the specific shape of glasses, as well as their spatial sequence in order to effectively remind them of what drinks to serve, and the order in which to serve them. Commenting on this work, Clark (2008) writes:

> *"The problem of remembering which drink to prepare next is thus transformed, as a result of learning within this prestructured niche, into the problem of perceiving the different shapes and associating each shape with a kind of drink. The bartender, by creating persisting spatially arrayed stand-ins for the drink orders, actively structures the local environment to press more utility from basic modes of visually cued action and recall. In this way, the exploitation of the physical situation allows relatively lightweight cognitive strategies to reap large rewards."* (pg.62)

Of course, individuals do not always structure their environment in ways that support the recall of information. Sometimes they deliberately structure their environment in ways that help them to forget. For example, people may avoid exposure to contexts and cues that remind them of unpleasant incidents. Such motivated changes in context or cue exposure can support forgetting processes in a number of ways (see Baddeley et al., 2009, chapter 10). Firstly, cue avoidance prevents memories from being retrieved, and the absence of retrieval practice makes the memory subject to normal decay processes. Secondly, changes in environmental context often bring about changes in incidental context (e.g. mood state), which then hinders the retrieval of unwanted memories. The main point here, as before, is that the individual can play an active role in structuring the environment so as to support the realization of a number of mnemonic objectives.

The World as External Memory

Our conventional view of memory is that it involves the storage of previously encountered information using the resources of our biological brains. Brain-based forms of encoding are obviously useful because we can rely on such encodings to be present whenever we need them to be so; for example, during the performance of specific tasks. However, there are situations in which the external environment may also provide reliable access to task-relevant information, and, in these cases, the world may be seen as serving as a form of external memory (O'Regan, 1992).

Recall the discussion about the process by which a spider constructs its web. In the context of that discussion it became apparent that the spider was using the current state of the environment (i.e. the structure of the web) as a means to structure and sequence its web-building behaviour. I suggested that the web effectively served as the spider's working memory of where it was in the process of web construction. The construct of working memory seems particularly apt here because working memory comes into play whenever we need to temporarily keep task-relevant information in mind for the purposes of accomplishing a specific task (e.g. as means of enabling us to progress through a series of problem-solving steps). Thus, one case where the use of working memory is evident is in the case of multiplying two numbers. Imagine that you are asked to solve a multiplication problem such as 21 x 3. If you are like me, you will solve this problem by multiplying 3 times 1, and then holding the result of this computation in mind, while simultaneously multiplying 3 times 2. Finally, you will combine the result of both computations yielding the answer – 63. This strategy of internal mental calculation and temporary storage is not, however, the only way in which we solve multiplication problems. An alternative strategy, particularly useful in the case of long multiplication, is to begin to draw on the representational resources of the external environment in order to help us with the cognitive demands of the task at hand. Thus, the strategy that most of us use in solving long multiplication problems is to draw on the resources of pencil and paper. We replace the sequence of neurally-encoded, imaginative manipulations of the problem-space with a series of perceptuo-motor routines that delegate important parts of the task-related representational burden to the external world. In place of purely inner computational operations we see a pattern of perception-action cycles in which single digit numbers are compared and intermediate computational results are stored in an external medium (i.e. the paper). This example, described in Wilson and Clark (2009), is a case of what we might call environmentally-extended computation or 'wide computationalism' (Wilson, 1994). It takes what is, ostensibly, an inner cognitive capability (an ability to do long multiplication) and shows how crucial aspects of the problem-solving process can be (and usually are) delegated to aspects of the external environment. What has changed in the transition to environmentally-based forms of long multiplication is that the use of neural resources to store information has been replaced with the resources of the external medium. Conceived of as a working memory task, the paper now appears to be playing the role of the visuo-spatial sketchpad component of working memory, while the human agent, in their interaction with the external resource, now appears to playing the role of the central executive. It seems, therefore, that physical action and environmental structure can sometimes serve the role of supporting working memory functions in a way that does not rely solely on internal forms of encoding. The essential idea is that, at least in some situations, the external world can function as a mnemonic resource that reduces the need to maintain an internal record of task-relevant information[4].

One question we can ask at this point is when does a human agent rely on a strategy of external representation and manipulation; at what point do they choose to switch from a strategy of inner, brain-based encoding to a strategy of real-world manipulation? The answer to this question seems to lie in the amount of time and effort associated with each strategy. Thus, if we think about the case of long multiplication it is likely that our decision regarding when to recruit the resources of pen and paper will depend on the amount of cognitive effort entailed by the multiplication challenge facing us. When the multiplication problem involves relatively small numbers (e.g. '3' and '21'), inner resources may be sufficient; however, as the demands increase with longer and longer numbers (e.g.

[4] This idea of the world serving as a form of external memory is not new. O'Regan (1992), for example, suggests that the world itself can sometimes function as form of external memory that can be continually interrogated in place an internal memory store. The idea is also apparent in more recent work concerning the nature of our visual conscious experiences (e.g. Myin & O'Regan, 2009), as well as in empirical work undertaken as part of the active vision paradigm (e.g. Ballard et al., 1997; see Findlay & Gilchrist, 2003).

'45367' and '23453'), we are increasingly inclined to distribute part of the information processing load to the external environment[5].

Efficiency and reduced cognitive load may be some of the reasons why a switch to externally-directed memory strategies is desirable; however, they are not necessarily the only ones. Sometimes, the real-world representation of certain information can support cognitive processing in a way that is not apparent when we rely on inner imaginings. Thus, consider Clark's (1997) discussion of the way in which academic papers typically get written:

> *"[the text] does not spring fully formed from inner cogitations. Instead, it is the product of a sustained and iterated sequence of interactions between my brain and a variety of external props. In these cases, I am willing to say, a good deal of actual thinking involves loops and circuits that run outside the head and through the local environment. Extended intellectual arguments and theses are almost always the products of brains acting in concert with multiple external resources. These resources enable us to pursue manipulations and juxtapositions of ideas and data that would quickly baffle the un-augmented brain." (pg. 207)*

In subsequent work, Clark (2007) writes:

> *"It is not always that fully-formed thoughts get committed to paper. Rather, the paper provides a medium in which, this time via some kind of coupled neural-scribbling-reading unfolding, we are enabled to explore ways of thinking that might otherwise be unavailable to us."*

The general idea here is that the act of physically representing information in the external environment and then reprocessing it via the medium of perception may deliver cognitive benefits that would not be seen in the case of purely internal processing. When it comes to memory, it may be that the physical representation and subsequent perception of retrieval cues serves as a much more effective means of prompting recall than the mere imagining of such cues. For example, when we hear a small piece of a song on the radio but cannot remember the whole song, it may be more effective for us to physically hum the part we heard rather than rely on inner rehearsal. Similarly, when we are asked to recall whole words from word fragments (as in 'M_M_R_Y') perhaps physical representations of the word fragments serve as a better aid to recall than simply imagining the fragment.

This notion of the relative efficacy of externally-represented cues as compared to purely imagined ones is still something that needs to be evaluated in empirical studies; however, there are some compelling insights into the value of external representation from the work on ambiguous images[6]

---

[5] The importance of effort as a deciding factor in switching to externally-directed processing routines is highlighted by the work of Gray and Fu (2004). They showed that when the cost (as measured in time) of accessing some external piece of information exceeds that required to access the same information from bio-memory, then subjects will resort to internal memorization strategies at the expense of externally-directed ones. It thus appears that the time taken to access information (what Gray and Fu refer to as information access cost) is the critical factor in determining the extent to which internal or external encodings are exploited during problem-solving activity. It is generally the case that we resort to using the environment as an external memory source whenever it is quicker and easier to do so.

[6] Ambiguous images are images which can be seen as either one of two different things. A famous example is the duck/rabbit image, which can be interpreted as the head of a duck or a rabbit.

(see Chambers & Reisberg, 1985). In one study, for example, subjects were asked to observe and remember an ambiguous image. They were then asked to seek an alternative interpretation for the image based on their mental image. Finally, the subjects were asked to draw the image and once again seek an alternative interpretation. The results from this study were surprising. Despite all the subjects reporting vivid mental representations of the original figure, none of them were able to discover the alternative interpretation of the figure from their mental images. In contrast, all of the subjects were able to see the alternative interpretation in their physical drawings (see Chambers & Reisberg, 1985). Commenting on these results Clark (2001) suggests that:

> *"...the subject's problem-solving capacities are significantly extended by the simple device of externalizing information (drawing the image from memory) and then confronting the external trace using on-line visual perception." (pg. 148)*

The processes by which this 'extension' in problem-solving capacities takes place is unknown at the present time. Further research is required to explore the phenomenon and examine its potential applicability to the case of physical cues in promoting better modes of retrieval performance.

Memory in Social Contexts

Extended mind accounts typically focus on physical props, aids and artefacts in describing how human cognitive processing is constituted by information processing networks that extend beyond the biological realm. Despite this, there is nothing in the bedrock claims of the extended mind thesis that precludes the possibility of socially-extended forms of cognition. Although some have expressed concern as to whether social resources could ever fulfil the criteria associated with extended minds (e.g. Clark & Chalmers, 1998), it is perfectly clear that, as a source of information, other people are perfectly able to fulfil the kind of action-guiding role played by (e.g.) notes, knots and other physical resources. In fact, some have even gone as far as to suggest that social resources are superior to physical resources in terms of fulfilling the conditions for cognitive extension. Tollefsen (2006) thus argues that the kind of limitations possessed by physical resources such as notebooks (for example, their inability to support the dynamic updating and revision of existing information (see Weiskopf, 2008)) are not applicable in cases where cognitive processing is socially distributed. In this case, the external memory resource (another person) is capable of the kind of dynamic updating capability that some commentators (e.g. Weiskopf, 2008) see as essential to cases of environmentally-extended memory.

When it comes to memory function, social forces and factors have not typically been viewed in the same light as their physical counterparts. New technologies, such as life-logging technologies, wearable sensor devices, global information networks, and semantically-enriched information resources have all been seen as heralding a new era in the technological augmentation of biological memory (Hodges et al., 2006; O'Hara, 2008; O'Hara et al., 2006b; O'Hara et al., 2009; Sellen et al., 2007). The same sense of optimism does not, however, seem to apply to the elements of our social networks. In this case, social influences on mnemonic functioning have typically been cast in a negative light. Social forces and factors have often been seen as constituting a negative influence on individual memory, with any form of social interaction potentially serving to corrupt or distort our recall of previously encountered information. As Barnier et al (2008) comment, "external influence is typically characterized as primarily negative, the relentless intrusion of the social into malleable individual memory."

Certainly, it is true that social influences can and do affect our memories, and sometimes the nature of this influence results in us remembering incorrectly (e.g. Loftus, 1997; Roediger et al., 2001). However, to denigrate all social influences as inherently malign is to overlook the important ways in which social relationships can structure and support memory-related activities. In particular, there are a number of ways in which the features of our social networks, and the nature of our social interactions within those networks, can and do work to exert a positive influence on mnemonic functioning (Barnier et al., 2008; Sutton, 2009; Sutton et al., 2010). In the sections below, I attempt to outline a number of ways in which social forces and factors can work to influence our access to the past.

Social Influences in Long-Term Memory Formation

Joint Reminiscences and Elaborative Rehearsal

When we think about the factors that determine whether or not we remember something over extended periods of time, we typically do not focus our attention on the features of our social networks. And yet the structure of our social networks, and the interactive exchanges that take place within them, play a potentially significant role in terms of influencing both the content of our memory and the form of our remembering activities. Consider, for example, a study by Barnier et al (2008), which found that people who maintained contact with old school friends had much better memories of their time at school compared to individuals who had lost touch with their school friends. There are clearly a number of ways in which we can account for this particular observation; however, one potentially fruitful line of investigation may be to see our memories of the past as being supported by the opportunities we have to revisit the past in the context of joint reminiscences. Thus, by virtue of the fact that we are able to talk about aspects of the shared past with our friends and colleagues, we are obliged to repeatedly recall those experiences, and this constitutes a form of socially-mediated rehearsal that serves to both strengthen our memories and make them more accessible on future occasions. When we fail to revisit the past in the context of joint discussions, our opportunities for rehearsal of those particular memories are reduced, and, in all likelihood, our access to those memories is weakened as a result. What this hopefully illustrates is that the content of our memories can be influenced by the features of our social networks. Social ties that endure for prolonged periods of time provide potent links to the distant past, and this can contribute to the formation of highly durable memories through the medium of joint reminiscences. In a sense, the nature of our social networks, and the nature of the communicative exchanges that take place within those networks, mediates our access to the past, and enables us to, in some sense, keep it alive.

But there is another aspect to this socially-mediated vivification of our past lives that goes beyond the simple idea of rehearsal. This is the idea that social conversation facilitates a specific form of rehearsal: one that is more suited to the formation of enduring long-term memories. Psychologists have obviously known that rehearsal plays an important role in memory consolidation for some time; however, it is now known that some types of rehearsal are better than others in supporting the formation of durable long-term memories. Craik and Lockhart (1972), for example, identified two kinds of rehearsal – maintenance rehearsal and elaborative rehearsal – in the context of their 'levels of processing theory'. Maintenance rehearsal is a kind of rehearsal that involves the simple repetition of to-be-remembered information, whereas elaborative rehearsal involves a much deeper kind of analysis: one in which the to-be-remembered information is often subjected to a more indepth semantic analysis. Importantly, Craik and Lockhart (1972) suggested that elaborative modes of rehearsal provide a more effective route to long-term memory as compared to the simple repetition of information (i.e. maintenance rehearsal).

Now, when we think about the way in which we access our memories in social contexts, and the way in which those memories are actually used to stimulate, maintain and shape the course of social interaction, it seems unlikely that the notion of maintenance rehearsal is one that adequately captures the dynamics of our joint reminiscences. Instead, what seems to be happening is that we are encouraged, stimulated and sometimes impelled to engage in a more effortful kind of analysis: a kind of analysis that, without the support of the social context, we might be reluctant (or find impossible) to engage in. So perhaps one way in which social exchanges contribute to memory is by effectively establishing the conditions under which we engage in more elaborative modes of rehearsal. By talking about a shared experience in a social context we are obliged to engage in a deeper kind of analysis of the past: exploring its significance for the actors involved, assessing its relationship to other experiences, and reflecting on its relevance to past, present and future goals. All this is likely to entail a deeper kind of analysis of the past than what would be undertaken in the case of solitary reminiscences. Even when we do reflect on the past by ourselves, it is unclear whether the nature of our processing activity is quite the same as that seen in social contexts. If nothing else, our social engagements typically require us to think about our recollections of the past from a different standpoint or perspective.

One reason why elaborative rehearsal may be more effective than maintenance rehearsal is because it supports the creation of more retrieval cues. Our ability to remember previously encountered information is influenced by a number of factors, including the associative strength between retrieval cues and the target memory, the relevance of retrieval cues and the number of retrieval cues available (see Baddeley et al., 2009, chapter 8). When we have more cues, we effectively have more routes to retrieval, and our chances of recalling the desired information are correspondingly increased. Perhaps something like cue creation happens in the course of our discussions about the past. Certainly it seems likely that in talking about a shared experience, we are required to examine it from a number of different perspectives and consider how it relates to other information. As part of this process we may come to associate a past event or situation with more linguistic cues, affective responses, and other associations. Typically, when we discuss events and experiences with others we do so in a way that establishes linkages and relationships to other information. This can include the emotional impact on ourselves and others, the reason why particular events and situations arose, and so on. By discussing and evaluating these relationships, it is possible to see our social interactions as supporting the realization of rehearsal processes that serve to consolidate otherwise fragile long-term memories. Since these processes extend across both the social and biological domains, it is possible to see them as a form of socially-extended cognition: one that works as part of our normal everyday mnemonic functioning.

Flashbulb Memories

Perhaps something like the aforementioned process of socially-mediated elaborative rehearsal can be used to account for the phenomenon of flashbulb memories (Brown & Kulik, 1977). Flashbulb memories are our memories of particularly dramatic and emotive events, such as the 9/11 attacks on the World Trade Center. These kind of memories are usually very vivid, and they tend to persist for a long period of time, which is surprising given that the events on which they are based tend to be of relatively short duration.

In order to explain flashbulb memories, Brown and Kulik (1977) proposed the existence of a separate kind of memory process that works to produce an almost photographic representation of the event in question. The process is presumed to be activated by stress or extreme emotion, which seems to be a prevailing feature of all the events associated with flashbulb memories.

Over the years, a number of theories have been developed to account for flashbulb memories. Joseph LeDoux (1996), for example, suggests that when the amygdala detects an aversive situation, it activates the autonomic nervous system, which in turn causes the release of adrenaline from the adrenal gland. It is this release of adrenaline that is implicated in the strengthening of explicit memories at the level of the temporal lobe memory system. In spite of this, not everyone is convinced that the phenomenon of flashbulb memory requires the existence of a separate type of memory process. Neisser (1997), for example, points out that the kind of events encoded by flashbulb memories are typically ones that we tend to discuss with others, and it is this discussion that makes flashbulb memories memorable, rather than any sort of emotion-mediated physiological encoding process. In support of this, Neisser (1997) recounts the results of a study undertaken in the aftermath of an earthquake, which shook northern California in 1989. The subjects in the study came from three different geographic regions, and they were each affected by the earthquake to a different extent. Interestingly, Neisser found that there were no significant correlations between emotional arousal and recall. In addition, the mean rated arousal of all the subjects in the study was not particularly high, a finding that Neisser attributed to the previous experience of Californian residents with earthquake incidents. Neisser (1997) subsequently asked what it was that made the earthquake so memorable for people close to the earthquake's epicentre. He writes:

> *"If the stress of the moment was not what made these experiences so memorable, the key factor must have been something that came into play afterward. In my view, that something made its appearance as soon as people realized that this had been no ordinary tremor. What did they do then? They started to talk about it. People who live through big events have many opportunities to describe their experiences, not just once, but over and over. They tell their stories to everyone they meet, and also to everyone who calls on the phone to see if they are all right. My hypothesis is that those rehearsals, the telling and retelling of each individual's earthquake narrative, are what made the experience of this particular earthquake especially memorable." (pg. 1701)*

What Neisser is suggesting here is that flashbulb memories occur for precisely the kind of reasons we have been discussing: people engage in conversation with other people, and this provides the kind of context in which optimal modes of mnemonic encoding are established.

It is important to bear in mind here that this sort of socially-centred theorizing does not necessarily undermine or downplay the importance of biological factors in terms of our understanding of memory phenomena. In the case of flashbulb memories, it may be perfectly true that the events we remember are those associated with strong levels of emotional arousal. And it may also be true that emotional arousal is playing an explanatorily significant role in our understanding of how flashbulb memories get formed. It may be, for example, that emotional arousal is still required to encourage people to seek out conversational opportunities and to make particular events the topic of protracted communicative exchanges. The point, for present purposes, is not that biological factors are irrelevant to our understanding of flashbulb memories; it is simply that we need to understand the kind of role they are playing in the cognitive processes associated with those memories. Thus, rather than emotional arousal playing an integral role in cognitive processing, what may be more likely, in the present case, is that emotion is serving as the instigating factor in a cognitive process that is *then* implemented in the social rather than the biological domain. Inasmuch as this is true, flashbulb memories may provide an example where it is the biological factors that serve as a causal influence to a cognitive process that is comprised of information processing loops that extend beyond the biological realm. Typically, in discussions of extended cognition, the reverse is claimed to be true. Some philosophers have thus argued that the mistake made by extended mind theorists is to move from the mere causal coupling of some object or process to a cognitive agent, to the claim that the object or process is actually part of the agent's cognitive apparatus. This has been dubbed the "coupling-constitution fallacy" (Adams & Aizawa, 2010), and it is typically reserved for cases where an external, non-biological factor is seen to exert a causal influence on neural processing. The case of flashbulb memories implies the same fallacy may be true when it comes to our understanding of the role played by biological factors in cases of environmentally-extended cognition. In this case, it is the biological factors that are playing a causal role (e.g. causing people to talk about certain events), and the cognitive process itself is constituted by information processing loops that draw on quite separate social and biological resources.

Influencing What We Forget

In addition to social relationships influencing our ability to remember particular events, it is also possible that they play an important role in determining what we forget. Whenever we talk about our past experiences with another person, we focus on some aspects of those experiences at the expense of others. Typically, we focus on those aspects of a situation that are personally relevant to us, or which are significant to the social context in which the past is discussed. This is important because research has shown that when we repeatedly recall some information at the expense of other, related information, then our access to the information that was not recalled is reduced. What seems to happen is that we effectively forget the non-recalled information.

One empirical demonstration of this is provided by the phenomenon of retrieval-induced forgetting (e.g. Anderson et al., 1994). In the retrieval-induced forgetting paradigm, participants are initially exposed to a set of category cue-exemplar word pairs. For example, they might be exposed to a list containing words pairs such as 'fruit-orange' and 'fruit-apple'. Subsequently, participants engage in repeated directed retrieval practice on half the exemplars from half of the category cues. For example, they might practice retrieving the 'fruit-orange' word pair, but not the 'fruit-apple' word pair. The way in which retrieval practice is performed is via category-stem recall, so participants are presented with a category-stem, such as 'fruit-or____' and asked to recall the exemplar word. In the test phase of the experiment, participants are presented with all the category cue words, and they are asked to retrieve as many of the exemplar words as possible. Results using this paradigm consistently reveal that participants' memory for exemplar words not associated with one of the category cues used in the second phase of the experiment is better than their memory for exemplar words that were associated with category cues but which were not recalled as part of the retrieval exercise. This is despite the fact that neither kind of word has been retrieved any more or less than the other. The phenomenon is referred to as retrieval-induced forgetting because it appears as though the repeated retrieval of particular words in response to a category cue has resulted in other words that share the same category cue being forgotten.

The phenomenon of retrieval-induced forgetting has also been observed in the context of autobiographical memories (Barnier et al., 2004). In one study, participants practiced the retrieval of some autobiographical memories at the expense of others, both of which shared the same category cue (for example, subjects might practice recalling one particular set of birthday memories rather than another set in response to the category cue of 'birthday'). The result was the standard retrieval-induced forgetting effect: a selective impairment of memories that were not practiced, but that were related to practiced memories via a shared category cue (see Barnier et al., 2004).

So one way in which the content of our memories might be influenced by our social relationships is via the selective discussion of particular events and experiences at the expense of other, related events and experiences. Imagine that a couple occasionally visit the same restaurant together. In this case, the restaurant serves as a category cue that can be associated with any number of specific episodic memories. The memory the couple have of their previous visits to the restaurant can now be influenced by the selective discussion of some particular visits at the expense of others. Perhaps one member of the couple will seek to recount the events of one particular visit for reasons that are particular to her, and this will serve as the basis for joint reminiscences of the evening together – the food that was eaten, the topics of conversation, and so on. This act of selective remembering not only makes the couple's memory of that particular evening more memorable, it also (according to the notion of retrieval-induced forgetting) makes the details of evenings that were not recalled less accessible. In this way, one or both members of the couple can effectively influence the memories that are jointly held by both members (with perhaps important implications for how each member of the couple thinks the relationship is progressing!). As Barnier et al (2004) note, the notion of retrieval-induced forgetting might have important implications for how we understand the way in which individuals actively construct their access to the past:

> *"The conditions modelled by RIF [retrieval-induced forgetting] might be one way that individuals shape their life story and manage unwanted memories. By rehearsing memories that support a positive rather than negative, ordered rather than disordered, life story, they may become highly accessible. Each time these memories are accessed they may reinforce the inhibition of competing, undesirable autobiographical knowledge." (pg. 473).*

What the present discussion adds to this notion is that our memories of the past may sometimes be the result of social forces and factors as opposed to merely individual concerns. The people that we interact with may structure our discussions about the past in such a way that the content of our memories says just as much about their interests, needs, and goals as it does our own. The past, in essence, may be the product of a delicate interplay between biologically-based processes on the one hand and the dynamics of socially-based information processing on the other.

Memories, Social Networks and the Self

The memories we hold about the past, both individually and collectively, are believed to play a crucial role in our notions of who and what we are. That is, our memories play an important role in determining our sense of personal identity. At the individual level, considerable attention has been devoted to an understanding of the bidirectional relationships between our memories and our sense of self. Conway (2005), for example, emphasizes the interconnectedness of self and memory in the context of his self-memory system. Conway proposes that our autobiographical memories contribute to our knowledge about who we are, and this in turn “constrains what the self is, has been and can be” (pg. 594.)

Inasmuch as our individual memories are the product of forces and factors that operate at the social level, it is possible to begin to see how social considerations play a role in influencing our sense of who we are. As with Conway’s theory, the relationship between self and the socially distributed processes that influence our memories are likely to be bidirectional in nature, with our social relationships influencing what we remember, and our sense of who we are influencing both the content and the form of the socially-distributed mnemonic processes that take place. What this leaves us with is the possibility that the mechanisms governing our sense of self are partly determined by the nature of our social interactions. By seeing at least some of our memory processes as extended to the social domain, we can see that social actors emerge as an important element of those forces and factors that help to make us what we are[7].

Collaborative Recall

Consider the following exchange (taken from Sutton et al (2010)) between a husband and wife discussing their honeymoon together, some 40 years previously. In this particular exchange, the couple are trying to recall the name of a show they both attended.

Wife: *And we went to two shows. Can you remember what they were called?*
Husband: *We did. One was a musical, or were they both? I don’t... no... one...*
W: *John Hanson was in it.*
H: *Desert Song.*
W: *Desert Song, that’s it, I couldn’t remember what it was called, but yes, I knew John Hanson was in it.*
H: *Yes.*

This discussion is quite typical of our attempts to recall some piece of shared information in social contexts, and it highlights something quite important about the nature of collaborative recall. Notice that the name of the show is successfully recalled as the result of cross-cuing between the couple. One person provides a cue, which by itself is inadequate to prompt the recall of desired information in either person. However, the cue does provide the basis for retrieval of another cue that is then fed back to the original person, and so on. This iterative cycle of engagement supports the progressive generation and elaboration of cues until eventually the conditions for one or other person to successfully recall the information are established. In this sense, it is the interaction between the individuals that provides the basis for recall. Absent the social interaction and the memory may be inaccessible. In fact, it may be largely impossible (or at least very difficult) for an individual, acting by themselves, to recall the details of a distant shared experience without the kind

[7] This may, of course, also extend to notions of collective memory and group identity. Inasmuch as our collective memories of shared experiences are determined by the dynamics of our social interaction, then our sense of who *we* are as a group – our group identity – may be influenced as much by socially-distributed memory processes as by intra-individual memory processes. Like the notion of personal identity, it is likely that the relationship between group identify and socially-extended memory is a bidirectional one: group identity influences the dynamics of socially-extended memory processes, and these in turn influence the collective memories that serve as the backdrop to group identity.

of mnemonic scaffolding provided by other people. In the example above, the couple initially only have very vague memories of the past event – the wife can remember the number of shows they went to, but the husband is unsure as to what kind of shows these were. Functioning by themselves, it seems unlikely that the recall of a specific piece of information – the name of one of the shows – would be very accessible. At the very least, each individual would need to generate a lot of additional retrieval cues, which may or may not be possible. And yet working together, the name of the show is successfully recalled within a very short space of time. Clearly, the couple, in this example, had shared memories of the past; however, access to those memories did not come about as a result of purely individual efforts – the couple worked together to establish the kind of informational conditions under which their respective bio-memories were most accessible.

What this example hopefully illustrates is that social interaction can sometimes play a productive role in enabling us to access specific memories. Just as when we repeatedly revisit aspects of the past through joint reminiscences, thereby engaging in elaborative rehearsal, social forces are here working to exert a positive influence on memory; they are supporting our access to the past in a way that might not be easily reproducible by other means. Without the kind of support provided by social interaction, many aspects of our past lives might forever be inaccessible to us. Dispense with long-standing social connections that have endured across many shared experiences together, and we potentially lose contact with parts our lives through which those connections have endured. Thus, while it might not seem that social networks have much to say about the content of our memories and the form of our remembering, we can see from this discussion that social networks can function in an important way to preserve access to our past lives. On this view, our social ties serve as more than bonds of social affiliation in the present, they also provide important linkages to our shared past together[8].

In the case of collaborative recall we see a process that is clearly social in nature, in the sense that it involves the cooperation of two or more human agents. Is this process a fitting candidate for a case of socially-extended cognition? One way of assessing this is to draw on what Clark and Chalmers (1998) refer to as the parity principle. The parity principle states that:

> *"...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 it is part of the cognitive system, then that part of the world is (for that time) part of the cognitive process." (Clark & Chalmers, 1998)*

The question we can now ask ourselves is if the process of collaborative recall were to take place in the head, would we regard this a cognitive process – as a process concerned with the retrieval of information stored in memory? On my view, this would be an acceptable conclusion. What we see in the case of collaborative recall are iterative cycles of information flow and influence that support the

[8] Perhaps this view of social bonds provides some insight into why the loss of an important friend or romantic partner is so personally significant. When long-standing social relationships are permanently broken, there is an important sense in which our access to past events and experiences is weakened. Our memories of those parts of the past we shared may no longer be as accessible as they once were, and, in the worst case, they may be forever irretrievable...essentially forgotten. Inasmuch as our memories make us who we are – inasmuch as they provide with a sense of our identity – then the dissolution of important social bonds and connections has potentially significant repercussions for our sense of who and what we are as individuals. When important friends and partners leave, an important part of who we are goes with them.

progressive creation, elaboration and structuring of cues in a way that ultimately leads to the recall of the target memory. As a form of cognitive process related to the retrieval of information from long-term memory, this seems perfectly acceptable, and it may even be a process that is not terribly far removed from what happens in the case of individual, biologically-based retrieval. What emerges from this discussion is a view of our social networks as potentially playing the same role in our memory processes as those implemented by the biological networks of the brain. Disrupt the social networks in some way and the consequences for memory may be akin to those seen in cases of neurological trauma.

The case of collaborative recall creates a picture of the mind that is empowered by its reliance on socially-distributed forms of information processing. At the same time, however, it also creates a picture of the mind that is perhaps more vulnerable than what we see in the case of individual (biologically-based) recall. The profligate extension of our cognitive machinery into the world seems to bring definite advantages, but it also seems to entail certain risks. In particular, the form of our cognitive processing now seems vulnerable to the vagaries of our social environments. The potential for disruption, distortion and deliberate subversion of our cognitive processing routines seems a distinct possibility, and it is not clear how well we can safeguard ourselves against these malign influences. This is something that takes us back to the beginning of this section and the concerns of those who see social influence as a potentially contaminative influence on memory. In the next section, I focus on a specific phenomenon that seems to call into question the value of social influence when it comes to our remembrances of the past.

Social Contagion

The cases of social influence discussed so far – the social mediation of long-term memory and the case of collaborative recall – portray a largely positive picture of how social forces and factors can contribute to human memory. In these cases, social processes were seen as enhancing or augmenting our mnemonic capabilities. However, not all forms of social influence are necessarily cast in such a positive light. One set of findings that motivates a more negative perspective of social influence comes from, so-called, social contagion or memory conformity studies (Gabbert et al., 2003; Meade & Roediger, 2002; Roediger et al., 2001). In these studies, the primary focus of interest is the way in which the presence of other individuals can affect the recall of previously encountered information. For example, in a study by Roediger et al (2001), people were placed in pairs (one of whom was a confederate of the experimenter). They were then asked to watch a slideshow depicting six common household scenes (e.g. a kitchen), each containing a variety of objects. In a subsequent recall phase, the two subjects were each asked to recall six items from the scenes that they had seen. However, the aim of the confederate at this stage of the experiment was to introduce items that had not been seen in the scenes. The issue addressed by Roediger et al (2001) was the extent to which the items introduced by the confederate would lead to the formation of false memories in the experimental subject. Would the experimental subjects, after hearing the items introduced by the confederate, incorporate these items into their own memory of the scenes they had witnessed? Roediger et al (2001) tested this in a third stage of the experiment. In this phase, the experimental subject was asked to recall as many items as possible from the six scenes without the presence of the confederate.

What Roediger et al (2001) discovered was that experimental subjects, exposed to the erroneous items during the collaborative recall phase, were more likely to recall these items in the final recall stage of the experiment. On the basis of these results, Roediger et al (2001) concluded that "subjects incorporated the confederates' memories into their own recollections"; the subjects' memories of the scenes they had witnessed had been effectively infected with the false memories suggested by the confederate.

At first blush, these experimental results seem consistent with a view that has become increasingly prevalent in the psychology of memory. This view emphasizes the vulnerability of our memories to the malign forces of social influence. As Elizabeth Loftus (1997) writes, "misinformation has the potential for invading our memories when we talk to other people" (pg. 51). The view thus regards social influence as compromising our ability to accurately recall past events and as potentially leading to the formation of false memories. The very name given to Roediger et al's experimental paradigm – social contagion – emphasizes the idea of social influence as something that is contaminative and subversive – of something that we should be highly circumspect about...or avoid altogether.

This negative perspective, however, is not necessarily vindicated by the results of false memory research. For a start, Roediger et al (2001) note that, in their experiment, experimental subjects would incorporate correct as well as incorrect items in their final recall list. In other words, when the confederate reported items that were actually present in the six scenes, experimental subjects showed a tendency to 'incorporate' these correct items into their memories as well as the incorrect ones. This particular finding does not suggest a socially-mediated disruption to recall capabilities, so much as a socially-mediated *improvement* in recall capabilities.

Thus, although the results of social contagion studies may be seen as revealing a notable shortcoming of our memories – namely their susceptibility to socially-mediated contamination – they may also reveal something of adaptive significance. Perhaps in many ecologically realistic situations (ones devoid of contrived, Machiavellian experimental setups) it does benefit us to make our memories susceptible to forms of social influence. In situations where we share our experiences with others, it may make little sense to overload our individual biological brains with the task of encoding all the details of a specific situation. Instead, it may make much more sense to distribute the encoding task across the elements of our social network – to share the burden of forming memories in the full knowledge that we can later rely on our social connections to facilitate our subsequent recall. If, in that situation, we discover that another's recollections of the past are different from our own, or if they recall aspects of a situation in more detail than we do, then it may make perfect sense for us to adjust our own memories in light of the information they provide. In this case, the susceptibility of our memories to social influence seems like a perfectly good thing: it enables us to check and, if necessary, modify our memories in light of the experiences of others.

This potentially adaptive role for socially-malleable memory is recognized by a number of writers, including those who work in the area of false memory research. Roediger et al (2001), for example, write that:

> *"If it is generally adaptive for people to update their memories based on other's recollections of the same events, then the occasional negative effects – creating false memories through erroneous reports – become understandable as an inevitable consequence of a normally adaptive process." (Roediger et al., 2001, pg. 370)*

Of course, none of this precludes the possibility that our socially-informed memories are necessarily accurate representations of the past. We can obviously be misled into forming false memories if the information provided by a social acquaintance is deliberately or accidentally fallacious. Two factors may guard against this in real-world situations. The first is that we may be more inclined to modify our memories when confronted with a greater number of consistent recollections. Thus, if one person's recollections of the past are at odds with the recollections of lots of other individuals, then it is possible that that person's recollections of the past are in error and require revision. In this case, it would make sense to converge on a common, consensual representation of the past. A second safeguard against false memories, this time resulting from attempts at social deception, may be to bias our susceptibility in favour of those we trust. Some empirical evidence in support of this has been provided by French et al (2008). Their results suggests that intimacy may be an important predictor of social contagion. They write that people are "even more susceptible to memory distortion when someone they know provides the misleading information" (pg. 271).

The preceding discussion touches on an important point in memory research. It is that once we take a given capability out of its proper social, ecological and evolutionary context, then many aspects of human memory can start to seem like design flaws or information processing shortcomings. This is especially so when we lose sight of the kind of features that may have characterized our ancestral physical and social environments, and against which our present-day cognitive capabilities may have taken shape. Take, for the instance, the idea, mooted above, that the social malleability of our memories may be an adaptive strategy to avoid the cognitive processing overhead associated with the detailed encoding of all aspects of an experienced event or situation. Such an approach seems like a sensible strategy for a social animal to pursue. Given limited cognitive resources, it benefits the individual to distribute the cognitive cost of encoding a situation and rely on a combination of relatively sparse individual encodings and the mechanisms of socially-distributed remembering (e.g. collaborative recall) in order to support later recall. If, during subsequent recall, an individual is exposed to new information about the situation that they did not encode, it is probably helpful, at that later stage, to incorporate the new information into their own memories of the situation.

Of course, one might argue that such claims are implausible for a number of reasons. Firstly, our social relationships seldom seem stable enough for us to delegate encoding and recall to them in the way suggested. Secondly, many of the situations where memory is called into action (e.g. revising for an exam) are not necessarily ones that we experience as a socially-shared activity. In many cases, activities that draw heavily on memory, like studying, are often insular ones that we undertake alone. Thirdly, to distribute encoding and recall to other people seems to make us unnecessarily vulnerable to deception and the vagaries of others' recall capabilities. These are obviously all valid concerns; however, they are based on the assumption that the context in which our memory capacities evolved resembles that in which we find ourselves today. This assumption is not necessarily valid. The social networks of our hominid ancestors 100,000 years ago may have been very different to those of today. In all likelihood, they would have been much more stable, with certain individuals providing a persistent social presence throughout the entire course of someone's life. In addition, it is likely that the kind of situations in which memory was called into service on the African savannah were profoundly different from those of the modern day. In all likelihood, situations that called for individual memory to function in isolation from the social realm would have been rare. Finally, regarding the risk of deliberate or accidental misinformation, this is an inevitable risk that comes with social malleable forms of memory. However, we can question whether either the risk of deception/misinformation or the severity of its consequences was as significant for our hominid ancestors as it apparently is in the present day. It may well have been the case, for example, that the stability of social networks encouraged individuals to become more trustworthy (at the very least it afforded the possibility of detecting and isolating those who could not be trusted). Furthermore, it is not necessarily clear what the consequences of misinformation would have been. Would fidelity of memory have been as important for our ancestors as it is today, or is the demand for accuracy something of relatively recent origin: something that appeared once we actually had a reliable means of checking our memories against pictorial and orthographic representations of the past (see below for more on this)? It is difficult to say for sure; however, fidelity is not necessarily the only selective pressure that sculpted our memory capabilities. The ability to form common, consensual representations of the past (collective memories) may have been important independent of the actual veracity of those representations. Such collective memories may fulfil a number of important social functions. For example, they may promote social cohesion and solidarity, and they may provide the basis for the formation of group identities. Inasmuch as this is true, then perhaps the relative benefit of converging on a common, collective representation of the past outweighed those associated with maintaining an accurate, but socially discrepant representation of past events and situations[9].

What I have tried to do in this section is show how a seemingly negative aspect of our mnemonic functioning – social contagion – can be reinterpreted in a more positive light once we begin to situate memory in its proper social, ecological and evolutionary context. The human mind probably evolved against a rich backdrop of social forces and factors, and not all of these forces and factors were necessarily benign (e.g. Byrne & Whiten, 1988). No doubt the human mind includes adaptations to protect itself from the exigencies of the social environment (e.g. the potential for social deception and manipulation); however, in all likelihood it also includes adaptations to *exploit* the features of those environments. One of those features is the presence of other minds, and the resources of those minds are things that can be factored deep into our own cognitive processing routines and information processing strategies. In general, evolutionary processes do not care about the material nature of the resources that are available to fulfil some function; they only care about how those resources can be exploited to serve adaptive outcomes. Artificial evolutionary processes attest to the variety of ways in which seemingly irrelevant forces and factors may be co-opted into a design solution. Thus, in using genetic algorithms to evolve real electronic circuits, Bird and Layzell (2002) managed to create an ‘oscillator circuit’ whose systemic oscillatory behaviour was parasitic on the radio signals being generated from a nearby computer. In essence, the evolving circuit had generated the correct oscillatory behaviour, but it had done so not by creating a genuine oscillator circuit; it had solved the problem by evolving radio reception capabilities and relaying the oscillations created by nearby circuits. Such phenomena are a common feature of many evolutionary processes. Thompson et al (1996) thus argue that during the evolution of electronic circuitry:

[9] In addition to social contagion, some of the other processes discussed in this section may also support convergence towards common, collective memories. Thus, in their discussion of how shared memories are created in conversation, Hirst and Echterhoff (2008) comment that socially shared retrieval-induced forgetting may be one of the means by which ‘individual memories converge onto a unified and stable rendering of the past’ (pg. 203). As mentioned above, such convergence may serve important social functions, such as supporting social cohesion and solidarity.

> *"…it can be expected that all the detailed physics of the hardware will be brought to bear on the problem at hand: time delays, parasitic capacitances cross-talk, meta-stability constraints and other low-level characteristics might all be used in generating the evolved behaviour." (pg. 21)*

What we begin to see, therefore, is that for any given purpose, evolution may often assemble solutions that pay scant regard to the actual material resources that are used to solve a particular problem. Inasmuch as the presence of others was a reliable element of our evolutionary past, then it is likely that many of our biologically-based processing routines will be sensitized to the exploitation of that presence. In this sense, it seems perfectly possible that the human brain would have opted for a strategy that relied on socially-distributed encodings and subsequent collaborative recall in order to alleviate itself of some of the cognitive load associated with the formation of highly accurate, detailed representations. The social contagion phenomenon may then be seen as a means by which our initially sparse encodings are supplemented with whatever details of a past event or situation are deemed significant enough (or important) enough to discuss. On this view, social contagion emerges as a perfectly normal feature of human memory, originating from our adaptive tendency to rely on the elements of our social network as a means of distributing the cognitive effort associated with mnemonic encoding.

Intermission: Lennie's Loss

Lennie, the main protagonist in the movie *Memento*, is, if you recall, an individual with a damaged mind. The injury sustained by Lennie during the attack on his wife means that he can no longer form new long-term memories. This leaves him with a number of impairments, some of which he attempts to remedy by relying on his system of hand-written notes, annotated photos and body tattoos. Lennie's bio-cognitive machinery is clearly damaged as a result of his injury, but what does this really mean in terms of the impairments that Lennie suffers? What exactly is it that Lennie has lost?

Perhaps most significantly, given Lennie's objectives in the movie, Lennie has lost an ability to efficiently piece together a temporally distributed body of facts in order to arrive at a correct understanding of some particular state-of-affairs. Lennie's main aim in the movie, remember, is to solve the mystery of his wife's murder, track down her killer and take his revenge. So, from Lennie's perspective, perhaps the most significant impact of his memory impairment relates to his inability to encode and retrieve facts relevant to the case. He has essentially lost an ability to 'connect the dots', to establish a network of relationships between otherwise disparate pieces of information in a way that progressively leads him to develop a better understanding of the mystery surrounding his wife's death.

In fact, however, there is no evidence from the movie that Lennie *does* see this particular kind of impairment as significant. Lennie believes in his system, and in fact he sees it as being more reliable than bio-memory. In addition, it is not clear that Lennie is in fact actually impaired when it comes to the assimilation, processing and utilization of facts. Lennie can store information, and he can use that information to guide his thought and action in goal-appropriate ways. A casual viewing of the movie might lead to the conclusion that, as the philosopher Joseph Levine (2009) sees it, "Leonard is totally clueless" (pg. 46). But to come to this conclusion is, I think, to miss something important about the movie. Lennie follows the facts of the case exactly where they lead him, and, at the end of the movie, they result in him killing the person whom he actually believes (justifiably so) to be his wife's killer. The notion that Lennie is "clueless" presumably stems from the fact that Lennie ends up[10] killing Teddy, someone who seems to have provided a degree of support and friendship to Lennie. Teddy is unlikely to have murdered Lennie's wife, so Lennie seems to have failed in his quest, and the roots of that failure seem to lie squarely at the feet of his system of external mnemonic artefacts.

Or does it? Do Lennie and his system really fail? We learn at the very end of the movie that things aren't quite as they had appeared to be. In fact, Lennie sets *himself* up to kill Teddy by manipulating the contents of his own external memory. Lennie relies on two things in order to accomplish this virtuoso display of self-deception. The first of these is that he accepts, without question, whatever information his external memory makes available to him. Just like the contents of his own bio-memory, Lennie trusts the contents of his external memory. The second thing that Lennie relies on is the fact that he won't be able to recall *why* information exists in his external memory. He won't, in other words, be able to remember the reason why he decided to store certain facts. Is it because those facts are important to the case, or is it because his past self has desires and intentions that are at odds with what his current self wants to do? Lennie just doesn't know. He can't appreciate the importance or the purpose of the information that is arrayed around him.

Both of these points highlight the significance of a key thematic element in *Memento*; in fact, maybe it is the most important theme of the whole movie. It is the theme of trust. Obviously, the nature of Lennie's injury makes him vulnerable to social manipulation and deception, but this vulnerability isn't necessarily an indication of the integrity of his external memory system. In fact, Lennie's system of storing and using external resources is relatively well-protected against at least some forms of social subversion and malicious tampering. For example, Lennie's recognition of his own handwriting serves to convince him that he is the originator of whatever is contained in his notes. Furthermore, it is difficult to see how Lennie could have failed to unequivocally endorse the information provided by his tattoos at the point they were created (especially since some of the tattoos were created by Lennie himself). It is also unlikely that the tattoos Lennie does have could easily have been edited without him being aware of it. As such, direct forms of social interference with the contents of Lennie's external memory are not necessarily a problem. What is a problem, however, is Lennie's ability to gauge the intentions of those around him. Practically everyone in the movie manipulates Lennie to their own ends. Lennie is acutely aware of this vulnerability – in fact, he has a tattoo on his lower left ribcage, which reads "DON'T TRUST". The irony here is that despite recognizing his vulnerability, there *is* someone who Lennie does trust implicitly: *himself*. Ultimately, this is the source of his undoing. The trust that Lennie places in himself, and the resulting faith he places in the facts that he has created, ultimately lead him to manipulate himself to his own ends.

So what emerges from the movie is a sense that Lennie's disability manifests itself in particular and perhaps unexpected ways. His ability to store and recall important factual information – his semantic memory, we might say – is preserved at least to some extent by virtue of his extended memory system. However, his ability to cope effectively with inter-personal relationships, and his ability to navigate the turbid, murky waters of the self is a deficit that is not remedied by the system he

[10] Because of the narrative structure of the movie, Lennie actually kills Teddy at the beginning of the movie. If the narrative elements of the movie were to be arranged in a chronologically linear sequence, the scene where Teddy is killed by Lennie would occur right at the end of the movie.

employs. In fact, it is precisely this feature of his extended memory that ultimately exacerbates his vulnerability to (social[11]) deception. "MEMORY IS TREACHERY", reads one of Lennie's tattoos. How true! It is probably the only one of Lennie's extended memories that is.

There is another aspect of Lennie's predicament that highlights the nature of his loss. During a particularly poignant scene in the movie, we find Lennie alone at night in a parking lot, huddled over a small fire into which he throws the few remaining items he has of his wife's belongings: a hairbrush, a well-worn book, and a teddy. This act of destruction reflects an attempt by Lennie to rid himself of the memory of his wife's loss. By destroying the last few mementos of his wife, he hopes to rid himself of the sense of grief he feels. Ultimately, of course, this is a pointless exercise. Lennie's inability to lay down new memories means that his most enduring memories are those associated with his wife...and her death. In the absence of any new memories, Lennie has no sense of the passage of time, and no ability to recover from the emotional trauma of his wife's loss. "How am I supposed to heal if I can't...feel time?", asks Lennie. This is Lennie's main problem, and it is the ultimate irony of the movie. We might initially be led to think that Lennie's loss is simply a matter of not being able to remember things. But we would be wrong. Lennie's main problem in the movie, the source of all of his pain, the reason for his murderous intent, and the basis for his, ultimately misdirected, sanguineous actions, lie not in his inability to remember; his main problem is his inability to forget.

Memory Technologies: Function, Features and Cognitive Extension

The increasing availability of wearable computing devices and ubiquitous information networks heralds new opportunities for the design and development of technologies to support human memory. Indeed, there are a number of technologies that are already being investigated for their memory-enhancing properties. These include wearable devices such as memory glasses to support semantic memory (DeVaul et al., 2005), sensor devices to support episodic memory (Berry et al., 2007; Hodges et al., 2006), and alerting devices to support prospective memory (Wilson et al., 2001)[12]. There are also efforts underway to explore the potential of life-logging technologies and semantic modes of information representation to support forms of autobiographical memory and the compilation of autobiographical knowledge bases (O'Hara et al., 2009). All of these technologies aim to capitalize on recent advances in electronic devices and the Web in order to improve aspects of human memory function.

When we think about technological innovation, it is common to think about the purpose or function of the technology. What should the technology do? What functions should it serve? And when it comes to the design and development of memory technologies, the answer to this question seems

[11] There is a side issue here concerning the nature of Lennie's act of self-deception. Does this constitute a true form of self-deception, or is it rather a case of social manipulation? Much depends here on whether we regard Lennie as the same person throughout the entire movie, or whether we regard Lennie as essentially a series of different people. These are questions of identity and they lie beyond the scope of the current paper. Nevertheless, on a personal level, it does feel as though the Lennie who wholeheartedly believes (albeit falsely) that Teddy is the killer is a different person from the Lennie who does not believe that Teddy is the killer, but who sets himself up to believe that such is the case in the future. Inasmuch as Lennie has a different identity at different points in the movie, it is easy to see how Lennie's act of self-deception could easily be construed as a case of social manipulation.

[12] These technologies obviously do not include a range of pharmacological interventions that could be used to enhance memory (see Lynch, 2002).

relatively straightforward: memory technologies should help us remember things better. Given our intuitions about what memory is, it seems relatively clear that memory technologies should function so as to improve our ability to recall past experiences and previously encountered facts. What else could such technologies possibly be used for?

Unfortunately, our intuitions about what memory technologies should do is complicated by a number of factors. In general, a new cognitive technology should aim to do one of three of things: 1) it should aim to redress or remedy an apparent shortcoming or limitation of an existing capability, 2) it should aim to augment an existing capability, or 3) it should aim to install new kinds of capabilities. Thus, when it comes to memory technologies, we might imagine them doing one or more of the following:

1. Addressing an apparent limitation of our existing mnemonic capabilities; for example, attenuating our tendency to forget things, or addressing our inability to keep more than a certain number of things in working memory at the same time.
2. Improving on our existing mnemonic capabilities, such as enabling us to recall more information with greater accuracy and fidelity.
3. Providing new kinds of mnemonic capability; for example, storing the totality of human knowledge in semantic memory.

The problem with this simple conception of what memory technologies should do is threefold. Firstly, psychologists have distinguished many different types of memory (e.g. semantic memory, episodic memory, autobiographical memory, procedural memory, prospective memory, and so on), and these types of memory differ (among other things) in terms of the kind of information that is retrieved and the purpose to which the retrieved information is put. As a result, it is difficult to come up with generalized statements as to the function of memory technologies. In general, memory technologies will target a specific kind of memory, and their functionality will vary accordingly. A second and potentially more significant problem is that it is not always clear that a deficiency or shortcoming in an aspect of our cognitive profile really is a weakness that needs to be remedied. Sometimes, what seems like an apparent shortcoming in cognitive processing can actually play an unanticipated adaptive role in our overall cognitive functioning, especially when such features are seen as operating with a broader developmental and social context. Thirdly, it is not necessarily clear that the true function of memory (as seen in either its historical, evolutionary or modern-day context) is one of accurate recall. While truth may be normative for some types of memory (e.g. semantic memory), this may not apply to all types of memory. In fact, when it comes to autobiographical memory, for example, a number of memory researchers have been keen to emphasize the primacy of other types of function; for example, social, psychodynamic and communicative functions.

In addition to these considerations, there is also the issue of whether a memory technology should aim to emulate aspects of bio-memory function or replicate bio-memory phenomena. This issue actually subsumes a perplexing range of issues, such as whether a memory technology should seek to emulate the representational and computational features of bio-memory (e.g. distributed, superpositional modes of information encoding) and whether it should exhibit the same kind of phenomena as those seen in the case of bio-memory (e.g. the generation effect).

In this section, I will attempt to answer at least some of these questions, or at least make them somewhat more intelligible. I first discuss the issue of correcting or addressing what seem like design flaws in our bio-memory systems. I focus, in particular, on our tendency to forget and the processing limitations of working memory. In the following section, I focus on the normativity of truth in memory technology design. Here I discuss to what extent memory technologies should support the accurate recall of previously encountered information. Finally, I explore the extent to which memory technologies should seek to emulate the features and characteristics of the bio-memory system.

Addressing Nature's Shortcomings

As a set of processes that enable us to recall information from the past, human memory shows an amazing mix of both strengths and weaknesses. Its strengths are evident whenever we are able to reflect on (enjoyable) events from the distant past, or whenever we are to recall with ease a set of well-rehearsed facts. It's weaknesses, however, are all too obvious. They include, most notably, the susceptibility of our memories to decay, our tendency to forget seemingly important pieces of information, the susceptibility of our memories to social contagion (Loftus, 1997; Roediger et al., 2001), and the processing limitations associated with some aspects of our memory system (e.g. working memory capacity). In light of these weaknesses, it seems plausible that one of the aims for memory technology development would be to address these weaknesses – to remedy the shortcomings that biology has bestowed upon us.

The problem with this development goal is that it is far from clear that the apparent weaknesses of bio-memory really are the kinds of things that should be the target of technological innovation. Take forgetting, for example. Our tendency to forget is undoubtedly a common source of frustration in many situations. However, to see forgetting as uniformly negative is to overlook its potentially adaptive role in enabling us to rid ourselves of irrelevant, redundant or damaging information. In recent years, human memory researchers have begun to appreciate the important, adaptive role of forgetting in our everyday behaviour. Forgetting may, for example, act as an updating mechanism that reduces the impact of past, irrelevant experience on current, relevant activity (Bjork, 1989). Forgetting may also enable us to avoid contact with aspects of the past that are psychologically and emotionally damaging (consider, for example, the way in which Lennie's inability to forget made it impossible for him to come to terms with his wife's loss). Thus, rather than see forgetting as evidence of a cognitive failure – as a shortcoming of a cognitive system that was designed to do something else – it may be worth considering the possibility that forgetting is just as much a functional part of our memory systems as our ability to remember things. Inasmuch as this is the case, we should perhaps be somewhat circumspect of technological interventions that strive to eliminate forgetting from our cognitive repertoire. Indeed, when it comes to technology design, theorists have often highlighted the potential value of forgetting mechanisms in managing the vast quantities of information made available by a combination of cheap computer memory and the advent of global information networks (Morris et al., 2006; O'Hara et al., 2006a). Forgetting capabilities might be especially useful if memory devices countenance information storage techniques that are based on discrete, symbolic encodings (as is currently the case with computer memory), as opposed to the distributed, superpositional encoding techniques of bio-memory. Symbolic modes of information storage typically require a greater number of representational resources to store new information, and this has implications for the amount of information that can ultimately be stored, as well the efficiency of the search and retrieval processes that operate over the stored information.

What about the notion that some aspects of our memory system suffer from information processing limitations that could be obviated by technological interventions? One potentially important target here is working memory, which has been shown to be restricted in terms of the amount of information it can handle at any particular point in time (Baddeley et al., 2009, chapter 3). We have already seen one way in which such processing constraints may be dealt with in the case of human agents solving long multiplication problems by resorting to pen and paper solutions. Are there perhaps other ways in which technologies could be developed to expand our working memory capacities across a greater variety of task domains?

Before we start to think about potential answers to this question, it is worth considering the implicit assumption on which this question is based. The assumption is that the processing limitations of working memory really do constitute something of a design limitation – some sort of functional shortcoming, overlooked by biological evolution, which is now in need of effective remediation by the trenchant ingenuity of technology designers. It is far from clear, however, that the processing characteristics of our working memory system really do represent some sort of functional shortcoming. Perhaps such characteristics play an adaptive role in our cognitive economy, but these roles have simply been overlooked by cognitive psychologists.

In order to get a better grasp on this sort of argument, it helps to consider the results of a study focusing on the learning abilities of artificial neural networks. The study in question concerns Jerry Elman's (1993) attempts to train an artificial neural network to process complex bodies of linguistic information. Elman (1993) was interested in whether a particular type of neural network, called a recurrent neural network (see Elman, 1990), could be trained to learn about aspects of grammatical structure, such as the ability to learn about verb agreement and clause embedding in the sentence 'The girls whom the teacher has picked for the play which will be produced next month practice every afternoon' (example from Elman, 1993). What Elman (1993) aimed to do was examine whether a recurrent neural network could learn about the grammatical structure of complex sentences, such as those exhibiting multiple clause embedding and long distance dependencies. In particular, the neural network was trained to take one word of a sentence at a time and predict what the next word in the sentence might be.

Unfortunately, Elman's efforts were in vain. The network completely failed to learn about the grammatical structure of the complex sentences. Not only did the network fail to develop a fully generalizable performance profile, it also failed to adequately master the data on which it was trained. In trying to account for these results, Elman (1993) tried an alternative training regime in which the network was initially presented with examples of very simple sentences and then exposed progressively to more complex ones. The aim was to isolate the precise point at which the network's performance broke down. At what level of sentential complexity would the network prove incapable of making further progress?

The results from this alternative training regimen were surprising. Elman (1993) discovered that when presented with staged training inputs (each increasing in complexity) the network was able to realize its original training objectives. Thus, what seemed to be important to a network's ultimate ability to learn about grammatically complex sentences was that its training was structured in such a way that it was able to learn about the simple cases first. Once the network was proficient in handling these simple cases, it was then able to learn to cope with progressively more complex ones.

It was almost as if the network's initial success with the simple cases laid the groundwork for subsequent learning about the more complex ones.

Moving on from this initial result, Elman went on to explore the effects of a further manipulation that built on the results of the staged input case. Rather than impose restrictions on the sequential order of exposure to training cases, Elman (1993) used an incremental memory solution in which the recurrent feedback provided by the layer of context units was gradually increased as training progressed. The effect of this manipulation was to limit the temporal window in which linguistic inputs could be processed. It thus forced the network to focus (at least initially) on the simplest training cases. And, as the memory provided by the recurrent units was increased over the course of training, so the network was able to deal with progressively more complex inputs. The effect of the incremental memory solution was thus the same as that achieved by staged training: it promoted an initial undersampling of the training data in such a way that the network's long-term ability to learn about complex regularities was enhanced. As Elman (1993) notes, this is an important discovery because it may shed light on the functional significance of a developmental progression in human cognitive capabilities. Thus, rather than see the working memory limitations of young children (see Kail, 1984) as a computational shortcoming that needs to be overcome in order to reveal the functional profile of adult cognition, Elman's (1993) findings suggest that such 'limitations' may play an important (and perhaps indispensable) role in children's cognitive development:

> *"Seen in this light, the early limitations on memory capacity assume a more positive character. One might have predicted that the more powerful the network, the greater its ability to learn a complex domain. However, this appears not always to be the case. If the domain is of sufficient complexity, and if there are abundant false solutions, then the opportunities for failure are great. What is required is some way to artificially constrain the solution space to just that region which contains the true solution. The initial memory limitations fulfil this role; they act as a filter on the input, and focus learning on just that subset of facts which lay the foundation for future success." (pg. 9-10)*

Elman's (1993) findings thus encourage us to think about the functional significance of what might otherwise be seen as a functional shortcoming. Seen from a particular perspective, it might seem as though working memory suffers from a number of performance limitations. However, such views largely ignore the potential role that such 'limitations' play in terms of the broader ecological (e.g. developmental and social) contexts in which a human being typically operates. Because most studies tend to look at particular populations of people (adults) performing in particular situations (laboratory-based studies), it is easy to see how interpretations and perspectives may be skewed. Once we isolate a cognitive system from its normal operating environment, many of its processing characteristics may appear mysterious, anomalous or ill-suited to the tasks that we present it with. In this situation, many aspects of a cognitive system may seem like design flaws. The moral of this discussion, therefore, is that we should not be too hasty in our efforts to identify and redress functional shortcomings. When seen in their proper ecological context, cognitive systems are likely to operate in complex ways with a variety of other forces and factors, and, in all likelihood, what initially seemed like a functional shortcoming may, on closer inspection, re-emerge as an essential feature of an environmentally-embedded and temporally-extended cognitive agent. Whenever we see a defect or deficiency in nature, a desirable feature is probably never far away.

Technologies for Total Recall

An oft maligned aspect of human memory concerns its susceptibility to error. In general, we want our memories of the past to be accurate, irrespective of whether we are recalling facts for an exam or whether we are trying to recall the details of last year's birthday party. Accuracy is also often highly prized in scientific studies of human memory. The golden yardstick against which our memory functions are measured in laboratory situations often concerns the correspondence of our memories to some previously encountered set of information items. Psychologists thus seek to 'test' human memory by measuring the extent to which we can recall information presented at some chronologically earlier stage of an experiment. Any errors in the recall process are typically seen as failures of human memory.

In light of all this, the goal of improving the accuracy of human memory seems like a legitimate and sensible target for new memory technologies. Few would question the value of technologies that enabled us to remember things with greater accuracy, and even fewer would perhaps suggest that memory technologies should encourage us to harbour false or incorrect representations of the past.

In spite of all this, however, the assumption that truth is normative for human memory is an assumption that should not go unchallenged. In general, we should not seek to develop technologies that augment or improve an existing cognitive capability without first understanding what the function of the underlying cognitive capability really is. It might seem that one of the goals of human bio-memory is the accurate recollection of previously encountered information, but we shouldn't necessarily accept this assumption without first examining the empirical evidence and subjecting the assumption to some critical questioning. Truth may be normative for some types of memory, but is this the case for all types of memory? And why should we assume that the biological function of human memory, in all its many forms, is actually about the veridical representation of the past? What is so great about truth, anyway?

Leonard Shelby, the main character in *Memento*, is clearly someone who, at first sight, appears to value truth. Most of his activities in the movie seem to be concerned with the accurate representation of facts as a means to guide his thoughts and actions in semantically-sensible ways. Ultimately, of course, we learn that this is only half the truth. Lennie is actually someone who manipulates facts in order to avoid certain truths; for example, the truth that his wife's death was a result of his own actions. It therefore turns out that Lennie's main goal is not the accurate representation of past reality; rather, he seeks to subvert reality in a way that benefits him. Lennie is not really a victim of his amnesia; he benefits from it: his impairment enables him to construct a version of events that is far more propitious than the events that actually transpired.

Following on from the example of Lennie, we can ask whether we would also be any better off if *all* our memories of the past were completely accurate. Perhaps, like Lennie, a certain amount of self-deception is not such a bad thing. As McKenna (2009) writes:

> *"It may well be that some degree of self-deception in human agency is a good thing. We are probably all a lot better off believing that we are a little better looking, smarter, more interesting, kinder, and more thoughtful than we really are. Were we to live in the constant light of the unvarnished truth about ourselves, we'd likely be so burdened that it would be hard to get up in the morning." (pg. 36)*

At the very least this line of questioning should give us pause for thought. There is nothing necessarily beneficial about remembering the past in all its glorious detail if all we end up with is an excess of socially and psychologically disabling recollections. Perhaps it helps if our memories are a little hazy and subject to error. And if they are going to be hazy, they might as well be hazy in the right way: in a way that portrays us in a positive light and bolsters our self-esteem.

This idea establishes contact with a body of research concerning the functions of human memory; in particular, autobiographical memory. Autobiographical memory refers to the memories we have of specific events across the lifespan, and it includes the body of personal knowledge (autobiographical knowledge) that is associated with our life experiences (Baddeley et al., 2009, see chapter 7). Autobiographical memory probably plays a key role in providing us with a sense of who we are as individuals (Conway, 2005), and for this reason the features and functions of autobiographical memory have been of considerable interest to those working in memory research.

Theories about the functions of autobiographical memory tend to fall into one of three categories (Alea & Bluck, 2003; Bluck, 2003; Pillemer, 2003; Wilson & Ross, 2003). These include, what Bluck (2003) describes as self (e.g. maintaining a largely favourable view of the self), social (e.g. supporting social interaction by providing the material for conversation) and directive (e.g. planning for present and future behaviour) functions. Of these, the self and social functions are the ones most likely to entail distortion, and, in fact, a variety of kinds of error have been documented in the context of autobiographical memory research. One kind of error involves people's judgements of when particular past events occurred. Wilson and Ross (2003) argue that people are motivated to push negative events further back in order to maintain a favourable view of themselves in the present. They show how people can distance themselves from negative events by pushing them into the distant past, so as to make them no longer relevant to the current self's well-being. In addition, positive events can be pulled forward in time so that the current self can continue to take credit for past successes. The function of these distortions, according to Wilson and Ross (2003), is to enable people to create and maintain a coherent – and largely positive – view of their present selves and associated circumstances.

There are, of course, occasions when accuracy in recall is important. These are typically cases when we are required to recall specific facts for particular purposes (e.g. examinations). In general, however, it is not clear that accuracy is the main function of memory, at least of our episodic memories. It is true that we tend to associate memory with our ability to learn, and in these cases accuracy seems quintessential to the development of new skills and abilities. But the purpose of memory is not to learn. Memory reflects our learning, but we learn little from our memories (if by learning we mean some sort of behavioural change). Whatever we learn from our experiences, it is not something that tends to get explicitly represented in consciousness. Rather, what gets represented in consciousness during our recollections of the past are the experiences themselves, as well as our affective and emotional responses to those experiences. There is no reason why such representations have to be entirely veridical, even if it were possible for them to be so.

When we think of our memories in an evolutionary context, there seem to be further reasons for doubting that accuracy is a central feature of episodic memory. In terms of our evolutionary history, it is has only been possible to check the veracity of our memories in recent times. Until physical recordings of the past could made, using orthographic and pictorial schemes, it was probably very difficult to gauge the accuracy of our memories. This suggests that the emphasis on accurate representation is something that may have emerged relatively recently in the history of our species. It is quite possible that our evolutionary forbears survived perfectly well with recollections of the past that were not entirely accurate, and, as mentioned above, there may very well have been a distinct advantage in allowing individual memories to converge on common, consensual representations. In this case, the benefits of mnemonic convergence at the collective level may have outweighed the costs of mnemonic distortion at the level of the individual (see above).

The purpose of this section has not been to negate or undermine the value of technologies that strive to improve the accuracy of human memory; there are clearly cases when such accuracy is an important concern. Nevertheless, I think it is important to be cautious of efforts that strive to provide blanket improvements in mnemonic accuracy across *all* types of memory. If accuracy is not at the functional heart of a particular kind of memory (e.g. autobiographical memory), then it is difficult to see how technological interventions that focus on accuracy are really contributing to an improvement or enhancement of that particular aspect of mnemonic functioning. In general, a technology improves a cognitive capability if it delivers a profile of enhanced performance in the same areas as that targeted by the original capability. If the technology works to improve performance in some other area, then the technology is not so much improving the existing capability as installing a new kind of capability, with all the attendant risks that that entails. If the primary function of our autobiographical memory system is to support self and social functions (for example, maintaining a positive view of the self, bolstering self-esteem, creating a coherent life story, and providing material to stimulate conversation), then it is unclear whether technologies that improve the accuracy of recall can actually serve to enhance autobiographical memory. Inasmuch as such technologies work in opposition to existing processes, they may in fact do more harm than good. Rather than focus on accurate recall, a much better focus of attention for such technologies might be to concentrate on how to support the human individual in constructing renditions of the past that enable them to function better in both the psychological and social domains. In some cases at least, this might mean that the best thing memory technologies could do would be to help us modify our recollections of the past...or even dispense with them altogether.

The Role of Biology in Memory Technology Design

An important issue in memory technology design concerns the extent to which candidate technologies are modelled on the representational and computational solutions countenanced by biology. This issue is particularly important in the current context because not everyone accepts that the kind of resources commonly encountered in discussions of the extended mind (hand-written notes, annotated photos, and so on) possess the kind of features that warrant their treatment as forms of extended memory. In the original treatment of the extended mind thesis, Clark and Chalmers (1998) emphasized the nature of the functional contributions made by both the internal (biological) and external (technological) resources. The outer resources counted as part of Otto's cognitive machinery, they suggested, because the functional role played the resources was sufficiently similar to the role played by inner resources as to warrant similarity of treatment: both kinds of resources played the same causal role in terms of guiding thought and action, and they should thus both be seen as legitimate parts of the machinery of cognition.

The problem when it comes to notions of extended memory, however, is that the differences between the biological and non-biological components – the brain on the one hand and notebook scribblings on the other – seem so profound as to negate any possibility of treating the relative contributions of the two components in the same way. The functional differences between the inner and outer resources seem so great that it seems utterly inappropriate to talk about them in the same cognitive terms. One critic of the extended mind thesis thus writes: "the external portions of extended 'memory' states (processes) differ so greatly from internal memories (the process of remembering) that they should be treated as distinct kinds" (Rupert, 2004, pg. 407).

One example of this sort of criticism is provided by Weizkopf (2008). Weizkopf suggests that a core feature of genuine beliefs is that they are informationally-integrated with one another. This means that as new information is received, all the agent's pre-existing beliefs are dynamically updated in order to ensure global consistency and coherence. Unfortunately, while this happens automatically in the case of the biological brain, it is not something that we see in the case of external encodings. In this case, the encodings are not subject to the sort of dynamic updating mechanism encountered in the biological realm, and thus they do not qualify as the material vehicles of belief states:

> *"To count as a belief a state has to be part of a system of states in which processes of integration and updating function to keep the subject's mental contents in epistemic equilibrium to some degree or other." (Weiskopf, 2008, pg. 268)*

To exemplify the differences between the inner and outer resources, Weiskopf suggests that a normal human agent who believes that MOMA is on 53rd street and who later learns that the museum has been demolished will no longer believe that the museum is on 53rd street – the original belief will be automatically revised upon receipt of the new information. In addition to this, other related beliefs (for example, the belief that MOMA is a good place to get a latte on 53rd street) will also be subject to this form of automatic updating mechanism. The result will be that, within a normal cognitive agent, all beliefs will be in a state of what Weiskopf calls "epistemic equilibrium". This is not so in the case of Otto – the mnemonically-impaired agent who relies on his notebook to store information. In this case, the inscriptions in the notebook are not subject to automatic updating, and this means, according to Weiskopf, that they should not count as the vehicles of Otto's dispositional belief states.

Inasmuch as Weiskopf is right, then, it seems that we do need to play close attention to the details of biological implementation. In particular, we need to understand what it is about the nature of biological encoding that makes bio-memory systems capable of the sort of automatic updating that Weiskopf describes. Before we accept this conclusion, however, it is worth considering a number of counterarguments. Firstly, just because informational integration and automatic updating is important (even if it is important), this does not mean that we have to emulate biological solutions in order to achieve the same result. The kind of capability highlighted by Weiskopf is something that could, in theory at least, be accomplished by technological components that are considerably more sophisticated than the case of a simple notebook. Thus, just because one particular kind of technological solutions lacks the requisite functional features Weiskopf alludes to, this does not mean that other kinds of more advanced technological solutions could not fit the bill, irrespective of whether or not they pay heed to biology.

Another caveat concerns the extent to which Weiskopf really is correct in his characterization of the way bio-memory actually works. As Clark (2005) comments, it is by no means clear that biological memory actually does manage to track and dynamically alter all beliefs in quite the way that

Weiskopf suggests. It seems entirely possible that logical consistency and global coherence of belief states is accomplished in some other way – perhaps via a kind of active reasoning process over bodies of related information that are associatively linked and simultaneously activated at retrieval time.

Finally, we can question whether Weiskopf is really right to place so much emphasis on the importance of informational integration and dynamic updating when it comes to establishing the conditions for cognitive extension. The emphasis is well-placed argues Weiskopf because 1) it identifies a key functional difference between the biological and non-biological components, and 2) the difference matters because informationally-integrated belief states support the kind of intelligent behaviours that we typically ascribe to cognitive agents. However, while it seems impossible to deny the first of these points, it is not clear that the second point necessarily follows from the first. Is it really the case that informational integration is a prerequisite for the kind of behaviours warranting the ascription of belief to an agent? Would we still not be inclined to explain behaviour in terms of belief states, irrespective of whether those beliefs were logically consistent with the other beliefs the agent held? Clark (2010) touches on this issue when he writes:

> *"True, that which is stored in Otto's notebook won't shift and alter while stored away. It won't participate in the ongoing underground reorganizations, interpolations, and creative mergers that characterize much of biological memory. But when called upon, its immediate contributions to Otto's behaviour still fit the profile of a stored belief. Information retrieved from the notebook will guide Otto's reasoning and behaviour in the same way as information retrieved from biomemory. The fact that WHAT is retrieved may be different is unimportant here."*

A more general response to concerns about functional differences between the biological and non-biological resources is to deny that any such differences are actually of importance when it comes to the realization of technologically-extended cognitive systems. The basis for this claim is that what matters when it comes to cognitive extension is not so much the functional differences between various resources as the way in which those resources support patterns of behaviour that warrant the ascription of particular cognitive states and processes to the agent in question. On this view, what it means for some technological component to be part of an extended memory system is that the component makes functional contributions to behaviour in ways that supports talk of that behaviour as being the product of particular memories and memory-related processes. It should be clear that there are a variety of ways in which this state-of-affairs could be brought about. In some cases, a memory technology might be required to emulate aspects of biological function, while in other cases some other type of implementation strategy might be called for. However, there is nothing that *requires* the outer components to function in precisely the same manner as the inner ones. In fact, in most cases, I suspect that the outer will not function like the inner simply because the inner and outer are required to function *together* in order to give rise to psychologically-interesting and interpretable patterns of behaviour. This emphasis on complementarity between the inner and outer resources has, in fact, been an increasingly common aspect of debates about the extended mind in recent years (Sutton, 2010).

None of this, of course, implies that a consideration of biology is irrelevant or unimportant when it comes to the implementation of biotechnologically-extended minds. Part of what it means to be a technology apt for incorporation into a biological agent's cognitive processing routines is that it be designed to work in concert with the agent's existing capabilities. This, then, is one reason why the study of biological systems is important: it provides an understanding of the kind of pre-existing capabilities that a candidate technology might build on, work with or exploit in order to yield desired behaviours.

Another reason for studying biological systems is, of course, the insight and inspiration they provide to complex engineering challenges (e.g. Shadbolt, 2004). In this case, the lessons learned from biology might provide productive inroads into a number of difficult capability areas for memory technology solutions. These include, among other things, the need for highly context-sensitive information retrieval solutions that are able to provide proactive support and guidance to an agent without the need for explicit instruction; the need for computationally efficient modes of information retrieval that are capable of dealing with vast repositories of data collected across the lifespan; and the need for information storage solutions that economize on the use of available representational resources.

Conclusion

Extended cognition accounts earn their scientific keep to the extent that they enable us to understand cognitive systems and cognitive phenomena. In this paper, I have discussed the notion of cognitive extension and applied it to a specific aspect of human cognition, namely memory. I have suggested that some kinds of memory phenomena may be better understood once we factor in the contributions made by the resources of our social and physical environments, and I have argued that in some cases, such as the case with socially-distributed remembering, there are compelling reasons for counting non-biological resources as part of an extended information processing network that constitutes the material basis of human memory. I thus conclude that an extended approach to memory research improves our understanding of existing research findings, and it also serves as a potent stimulus for further scientific research.

The possibility of extended memory systems heralds rich opportunities for technology design and development. In particular, if the machinery of memory is partly constituted by information processing loops that extend beyond the neural realm, then it seems we have an easily accessible means by which memory performance can be influenced. By targeting the non-neural elements of the memory system, technological interventions could enhance memory in ways that are just as profound as those promised by the vision of memory-enhancing drugs. However, before we wholeheartedly embrace the vision of bio-technologically extended memory, it is important to understand exactly what the functional target of such technologies should be. In this paper, I have argued that a simplistic notion of memory technologies as promoting the recall of greater and greater amounts of information with ever-increasing degrees of accuracy may not be warranted. The design of memory technologies should be based on a careful consideration of what it is that memory buys us. In some cases, memory enables us to relive the past, while in other cases it enables us to be relieved of the past. The design of memory technologies should thus be based on a better understanding of the memory system as situated in its proper ecological context – that extended nexus of representational and computational elements that helps to make us who we are.

References

Adams, F., & Aizawa, K. (2010) Defending the bounds of cognition. In R. Menary (Ed.), *The Extended Mind*. MIT Press, Cambridge, Massachusetts, USA.

Alea, N., & Bluck, S. (2003) Why are you telling me that? A conceptual model of the social function of autobiographical memory. *Memory*, 11(2), 165-178.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994) Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 1063-1087.

Baddeley, A., Eysenck, M. W., & Anderson, M. C. (2009) *Memory*. Psychology Press, Hove, East Sussex, UK.

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997) Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4), 723-742.

Barnier, A., Hung, L., & Conway, M. (2004) Retrieval-induced forgetting of emotional and unemotional autobiographical memories. *Cognition & Emotion*, 18(4), 457-477.

Barnier, A. J., Sutton, J., Harris, C. B., & Wilson, R. A. (2008) A conceptual and empirical framework for the social distribution of cognition: The case of memory. *Cognitive Systems Research*, 9(1-2), 33-51.

Barrett, L. (2011) *Beyond the Brain: How Body and Environment Shape Animal and Human Minds*. Princeton University Press, Princeton, New Jersey, USA.

Beach, K. (1988) The role of external mnemonic symbols in acquiring an occupation. In M. M. Gruneberg & R. N. Sykes (Eds.), *Practical Aspects of Memory* (pp. 342-346). Wiley, New York.

Berry, E., Kapur, N., Williams, L., Hodges, S., Watson, P., Smyth, G., Srinivasan, J., Smith, R., Wilson, B., & Wood, K. (2007) The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: A preliminary report. *Neuropsychological Rehabilitation*, 17(4), 582-601.

Bird, J., & Layzell, P. (2002) *The evolved radio and its implications for modelling the evolution of novel sensors*. Congress on Evolutionary Computation (CEC'02), Washington D.C., USA.

Bjork, R. A. (1989) Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*. Erlbaum, Hillsdale, New Jersey, USA.

Bluck, S. (2003) Autobiographical memory: Exploring its functions in everyday life. *Memory*, 11(2), 113-123.

Brooks, R. A. (1991) Intelligence without representation. *Artificial Intelligence*, 47, 139-160.

Brown, R., & Kulik, J. (1977) Flashbulb memories. *Cognition*, 5, 73-99.

Byrne, R. W., & Whiten, A. (1988) *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press, Oxford, UK.

Chambers, D., & Reisberg, D. (1985) Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, 11(3), 317-328.

Clark, A. (1997) *Being There: Putting Brain, Body and World Together Again*. MIT Press, Cambridge, Massachusetts, USA.

Clark, A. (2001) *Mindware: An Introduction to the Philosophy of Cognitive Science*. Oxford University Press, Oxford, UK.

Clark, A. (2005) Intrinsic content, active memory and the extended mind. *Analysis*, 65(285), 1-11.

Clark, A. (2007) Curing Cognitive Hiccups: A Defense of The Extended Mind. *Journal of Philosophy*, 104(4), 163-192.

Clark, A. (2008) *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press, New York, USA.

Clark, A. (2010) Memento's revenge: the extended mind, extended. In R. Menary (Ed.), *The Extended Mind*. MIT Press, Cambridge, Massachusetts, USA.

Clark, A., & Chalmers, D. (1998) The Extended Mind. *Analysis*, 58(1), 7-19.

Conway, M. (2005) Memory and the self. *Journal of Memory and Language*, 53(4), 594-628.

Craik, F. I. M., & Lockhart, R. S. (1972) Levels of processing: a framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, 11, 671-684.

Dawson, M. R., Dupuis, B., & Wilson, M. (2010) *From Bricks to Brains: The Embodied Cognitive Science of LEGO Robots*. AU Press, Edmonton, Alberta, Canada.

Dennett, D. C. (1996) *Kinds of Minds: Toward an Understanding of Consciousness*. Basic Books, New York, USA.

DeVaul, R., Pentland, A., & Corey, V. (2005) *The memory glasses: subliminal vs. overt memory support with imperfect information*. 7th IEEE International Symposium on Wearable Computers, White Plains, New York, USA.

Elman, J. (1990) Finding structure in time. *Cognitive Science*, 14, 179-211.

Elman, J. L. (1993) Learning and development in neural networks: The importance of starting small. *Cognition*, 48(1), 71-99.

Findlay, J. M., & Gilchrist, I. D. (2003) *Active Vision: The Psychology of Looking and Seeing*. Oxford University Press, Oxford, UK.

Foelix, R. F. (1996) *Biology of Spiders*. Oxford University Press, New York, USA.

French, L., Garry, M., & Mori, K. (2008) You say tomato? Collaborative remembering leads to more false memories for intimate couples than for strangers. *Memory*, 16(3), 262-273.

Gabbert, F., Memon, A., & Allan, K. (2003) Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology*, 17(5), 533-543.

Gray, W. D., & Fu, W. T. (2004) Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science*, 28(3), 359-382.

Haugeland, J. (1998) Mind embodied and embedded. In J. Haugeland (Ed.), *Having thought: Essays in the Metaphysics of Mind*. Harvard University Press, Cambridge, Massachusetts, USA.

Hirst, W., & Echterhoff, G. (2008) Creating shared memories in conversation: toward a psychology of collective memory. *Social Research*, 75(1), 183-216.

Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., & Wood, K. (2006) *SenseCam: A retrospective memory aid*. 8th International Conference on Ubiquitous Computing, Orange County, California, USA.

Hurley, S. (1998) *Consciousness in Action*. Harvard University Press, Massachusetts, USA.

Kail, R. (1984) *The Development of Memory*. W. H. Freeman, New York, New York, USA.

Krink, T., & Vollrath, F. (1997) Analysing Spider Web-building Behaviour with Rule-based Simulations and Genetic Algorithms. *Journal of Theoretical Biology*, 185(3), 321-331.

Krink, T., & Vollrath, F. (1998) Emergent properties in the behaviour of a virtual spider robot. *Proceedings of the Royal Society of London B: Biological Sciences*, 265(1410), 2051-2055.

Krink, T., & Vollrath, F. (1999) Virtual Spiders Guide Robotic Control Design. *IEEE Intelligent Systems*, 14(5), 77-84.

LeDoux, J. (1996) *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon & Schuster Inc., New York, New York, USA.

Levine, J. (2009) Leonard's System: Why It Doesn't Work. In A. Kania (Ed.), *Memento: Philosophers on Film*. Routledge, Abingdon, Oxon, UK.

Loftus, E. (1997) Creating false memories. *Scientific American*, 277(3), 70-75.

Lynch, G. (2002) Memory enhancement: the search for mechanism-based drugs. *Nature Neuroscience*, 5(supp), 1035-1038.

McKenna, M. (2009) Moral Monster or Responsible Person? Memento's Leonard as a Case Study in Defective Agency. In A. Kania (Ed.), *Memento: Philosophers on Film*. Routledge, Abingdon, Oxon, UK.

Meade, M. L., & Roediger, H. L. (2002) Explorations in the social contagion of memory. *Memory & Cognition*, 30(7), 995-1009.

Morris, R., Hitch, G., Graham, K., & Bussey, T. (2006) Learning and Memory. In R. Morris, L. Trassenko & M. Kenward (Eds.), *Cognitive Systems: Information Processing Meets Brain Science*. Elsevier Academic Press, London, UK.

Myin, E., & O'Regan, J. K. (2009) Situated perception and sensation in vision and other modalities: a sensorimotor approach. In M. Ayded & P. Robbins (Eds.), *Cambridge Handbook of Situated Cognition*. Cambridge University Press, Cambridge, UK.

Neisser, U. (1997) The ecological study of memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352(1362), 1697-1701.

O'Hara, K. (2008) *The technology of collective memory and the normativity of truth*. Workshop on Philosophy and Engineering, Royal Academy of Engineering, London, UK.

O'Hara, K., Hall, W., van Rijsbergen, K., & Shadbolt, N. (2006a) Memory, reasoning and learning. In R. Morris, L. Trassenko & M. Kenward (Eds.), *Cognitive Systems: Information Processing Meets Brain Science*. Elsevier Academic Press, London, UK.

O'Hara, K., Morris, R., Shadbolt, N., Hitch, G., Hall, W., & Beagrie, N. (2006b) Memories for life: a review of the science and technology. *Journal of The Royal Society Interface*, 3(8), 351-365.

O'Hara, K., Tuffield, M., & Shadbolt, N. (2009) Lifelogging: Privacy and Empowerment with Memories for Life. *Identity in the Information Society*, 1(1), 155-172.

O'Regan, J. K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46(3), 461-488.

Pfeifer, R., & Bongard, J. (2007) *How the Body Shapes the Way We Think: A New View of Intelligence*. MIT Press, Cambridge, Massachusetts, USA.

Pfeifer, R., & Scheier, C. (1998) Representation in natural and artificial agents: an embodied cognitive science perspective. *Zeitschrift für Naturforschung*, 53(7-8), 480-503.

Pillemer, D. (2003) Directive functions of autobiographical memory: The guiding power of the specific episode. *Memory*, 11(2), 193-202.

Robbins, P., & Ayded, M. (2009) *The Cambridge Handbook of Situated Cognition*. Cambridge University Press, New York, USA.

Roediger, H., Meade, M., & Bergman, E. (2001) Social contagion of memory. *Psychonomic Bulletin and Review*, 8(2), 365-371.

Rowlands, M. (2006) *Body Language: Representation in Action*. MIT Press, Cambridge, Massachusetts, USA.

Rupert, R. (2004) Challenges to the Hypothesis of Extended Cognition. *Journal of Philosophy*, 101(8), 389-428.

Sellen, A., Fogg, A., Aitken, M., Hodges, S., Rother, C., & Wood, K. (2007) *Do life-logging technologies support memory for the past?: An experimental study using sensecam*. SIGCHI Conference on Human Factors in Computing Systems, San Jose, California, USA.

Shadbolt, N. (2004) Nature-inspired computing. *IEEE Intelligent Systems*, 19(1), 2-3.

Smart, P. R. (2010) *Cognition and the Web*. 1st ITA Workshop on Network-Enabled Cognition: The Contribution of Social and Technological Networks to Human Cognition, Maryland, USA.

Smart, P. R., Engelbrecht, P. C., Braines, D., Strub, M., & Giammanco, C. (2010) The Network-Extended Mind. In D. Verma (Ed.), *Network Science for Military Coalition Operations: Information Extraction and Interaction*. IGI Global, Hershey, Pennsylvania, USA.

Sutton, J. (2009) Remembering. In M. Ayded & P. Robbins (Eds.), *Cambridge Handbook of Situated Cognition*. Cambridge University Press, Cambridge, UK.

Sutton, J. (2010) Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (Ed.), *The Extended Mind*. MIT Press, Cambridge, Massachusetts, USA.

Sutton, J., Harris, C. B., Keil, P. G., & Barnier, A. J. (2010) The psychology of memory, extended cognition, and socially distributed remembering. *Phenomenology and the Cognitive Sciences*, 1-40.

Tarsitano, M., & Jackson, R. (1997) Araneophagic jumping spiders discriminate between detour routes that do and do not lead to prey. *Animal Behaviour*, 53(2), 257-266.

Thompson, A., Harvey, I., & Husbands, P. (1996) Unconstrained Evolution and Hard Consequences. In E. Sanchez & M. Tomassini (Eds.), *Toward Evolvable Hardware: The Evolutionary Engineering Approach* (pp. 136-165). Springer-Verlag, Berlin, Germany.

Tollefsen, D. P. (2006) From extended mind to collective mind. *Cognitive Systems Research*, 7(2-3), 140-150.

Triantafyllou, M., & Triantafyllou, G. (1995) An efficient swimming machine. *Scientific American*, 272(3), 64-71.

Weiskopf, D. (2008) Patrolling the Mind's Boundaries. *Erkenntnis*, 68(2), 265-276.

Wheeler, M. (2005) *Reconstructing the Cognitive World*. MIT Press, Cambridge, Massachusetts, USA.

Wheeler, M., & Clark, A. (1999) Genic representation: reconciling content and causal complexity. *British Journal For The Philosophy Of Science*, 50(1), 103-135.

Wilcox, S., & Jackson, R. (2002) Jumping Spider Tricksters: Deceit, Predation, and Cognition. In M. Bekoff, C. Allen & G. M. Burghardt (Eds.), *The Cognitive Animal*. MIT Press, Massachusetts, USA.

Wilson, A., & Ross, M. (2003) The identity function of autobiographical memory: Time is on our side. *Memory*, 11(2), 137-149.

Wilson, B. A., Emslie, H. C., Quirk, K., & Evans, J. J. (2001) Reducing everyday memory and planning problems by means of a paging system: A randomised control crossover study. *Journal Of Neurology, Neurosurgery And Psychiatry*, 70, 477-482.

Wilson, M. (2002) Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625-636.

Wilson, R. A. (1994) Wide computationalism. *Mind*, 103(411), 351-372.

Wilson, R. A. (2000) The mind beyond itself. In D. Sperber (Ed.), *Misrepresentations: A Multidisciplinary Perspective* (pp. 31-52). Oxford University Press, New York, USA.

Wilson, R. A. (2004) *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. Cambridge University Press, New York, USA.

Wilson, R. A., & Clark, A. (2009) Situated Cognition: Letting Nature Take its Course. In M. Ayded & P. Robbins (Eds.), *Cambridge Handbook of Situated Cognition*. Cambridge University Press, Cambridge, UK.