Human-Extended Machine Cognition

Paul R. Smart[a,*]

[a]Electronics & Computer Science, University of Southampton,
Southampton, SO17 1BJ, UK

Abstract

Human-extended machine cognition is a specific form of artificial intelligence in which the casually-active physical vehicles of machine-based cognitive states and processes include one or more human agents. Human-extended machine cognition is thus a specific form of extended cognition that sees human agents as constituent parts of the physical fabric that realizes machine-based cognitive capabilities. This idea is important, not just because of its impact on current philosophical debates about the extended character of human cognition, but also because it helps to focus scientific attention on the potential role of the human social environment in realizing novel forms of artificial intelligence. The present paper provides an overview of human-extended machine cognition and situates the concept within the broader theoretical framework of active externalism. The paper additionally shows how the concept of human-extended machine cognition can be applied to existing forms of human–machine interaction, especially those that occur in the context of the contemporary Internet and Web.

Keywords: Extended Cognition, Active Externalism, Artificial Intelligence, Cognitive Computing, Human–Computer Interaction, Human Computation

1. Introduction

In the attempt to understand the mechanisms of human intelligence, the human brain is a structure of central significance. In fact, given the way in which

---
*Corresponding author
Email address: ps02v@ecs.soton.ac.uk (Paul R. Smart)

cognitive phenomena are seen to emerge from the whirrings and grindings of the neurological machinery, it might be assumed that all that matters to human intelligence (from the standpoint of mechanistic realization, at least) is to be found solely within the neural realm—that the point source of intelligent thought and action is located inside the heads of human agents. This view, which I will dub the *neurocentric view*, sees the biological brain as the sole realization base for the human mind. According to the neurocentric view, human mental states and processes are the direct product of what the brain does. The human brain, in other words, is that part of the material world that realizes all humanly-relevant cognitive phenomena.[1]

An alternative vision of the mechanistic underpinnings of human intelligence comes in the form of a philosophical position known as active externalism (Clark, 2008; Clark & Chalmers, 1998). In contrast to the neurocentric view, advocates of active externalism propose that the machinery of the human mind is not restricted to the inner sanctum of the neural realm. Instead, they claim that the causally-active physical vehicles of the mind can, on occasion, extend beyond the traditional biological borders of skin and skull to include a range of non-neural (and even non-biological) elements. Active externalism thus provides us with an *extended view* of human cognition—a view in which the physical machinery of the mind is occasionally able to escape its cranial confines and extend out into the world.

Active externalism is a philosophical position of considerable interest and importance to cognitive science. In our quest to understand the material bases of intelligent thought and action, it is obviously important that we focus our attention on those parts of the physical world that are most likely to contain the mechanisms that are responsible for phenomena of interest (Craver & Tabery, 2016). If our attention is focused on only one part of what is, in effect, a

---

[1]Harnad and Dror (2006) thus embrace the neurocentric view when they suggest that "cognition takes place entirely within the brains of cognizers...cognition begins and ends at the cognizer's sensor and effector surfaces" (p. 9).

larger mechanistically-relevant matrix, then we face the risk that an important array of explanatorily-relevant forces and factors will end up falling beyond our field of view. The result is that a commitment to active externalism implies a shift in scientific focus. For inasmuch as we accept the basic tenets of the active externalist position, then it seems that our attempt to understand human intelligence will need to focus on more than just the biological brain. This looks to be particularly important if it is the porosity of the human cognitive system—i.e., our capacity to assimilate extra-organismic resources deep into our cognitive routines—that lies at the heart of our species' peculiar and prodigious form of cognitive success.[2]

But it is not just our approach to human intelligence that is affected by claims about the extended character of human cognition; active externalism is also a philosophical position that is of substantive relevance to the field of Artificial Intelligence (AI). In one sense, of course, this is trivially true. If the profile of human cognition is one that relies on a capacity for cognitive extension (i.e., a capacity to assimilate extra-organismic elements into cognitive processing routines), then attention to the details of that capacity is likely to yield rewards in terms of the attempt to engineer systems that seek to emulate (or surpass) human capabilities. There is, however, also a sense in which active externalism may be relevant to AI irrespective of whether or not we accept claims about the extended character of human cognition. It may be the case, for example, that an extended perspective helps us see how an array of difficult or intractable computational problems could be solved by exploiting cognitively-potent forms of causal commerce between an array of materially-heterogeneous resources. By combining the distinctive representational and computational capabilities of resources drawn from both the biological and the technological realm it may thus be possible to yield solutions that reduce (e.g.) the temporal and energetic costs

---

[2]Clark (2003), for example, notes that "what is special about human brains, and what best explains the distinctive features of human intelligence, is precisely their ability to enter into deep and complex relationships with non-biological, props, and aids" (p. 5).

associated with the performance of certain kinds of cognitive task (see Jonker, 2008). In this sense, an extended approach to machine intelligence dovetails with work that emphasizes the value of situated, embodied and enactive approaches to the design of AI systems (Chrisley, 2003; Froese & Ziemke, 2009; Iida et al., 2004; Lindblom & Ziemke, 2003).

In the present paper, I attempt to extend the conventional focus of active externalist theorizing by outlining an extended approach to AI.[3] In particular, I claim that when we look at the environment in which specific instances of machine intelligence are situated, we sometimes encounter a state-of-affairs in which one or more human agents serve as part of the physical fabric that realizes episodes of machine-based cognizing. This idea is captured by the thesis of Human-Extended Machine Cognition (HEMC):

> **Thesis of Human-Extended Machine Cognition**
>
> The casually-active physical vehicles of machine-based cognitive states and processes may, on occasion, include one or more human agents. Human agents are thus candidate parts of the (extended) realization base for machine-based cognitive capabilities.

The HEMC concept, it should be clear, reverses the usual focus of active externalist theorizing. Instead of the idea that (e.g.) a technological device serves as part of the material fabric that realizes a specific instance of human cognizing, the HEMC thesis encourages us to switch our point of view and look at things from the device's 'perspective'. In this case, the nature of the coupling between the device and the human user can be seen to provide opportunities for *bidirectional* forms of cognitive extension. If we accept the idea that a

---

[3]The application of active externalist theorizing to AI systems has been the focus of previous work. Jonker (2008), for example, provides a useful overview of some of the potential advantages associated with cognitive extension in the case of AI systems. The focus of the present paper is on a specific form of cognitive extension that involves the incorporation of *human* agents into machine-based cognitive processing routines. This is not something that is explicitly addressed by Jonker.

technological artifact may, on occasion, form part of the supervenience base for human cognitive states and processes, then it is surely possible, at least in principle, for the human agent to also, on occasion, form part of the material fabric that realizes the cognitive states and processes of a technological artifact.

The main aim of the present paper is to introduce the HEMC concept and situate it within the broader philosophical and cognitive scientific literature. I also attempt to show how the HEMC concept can be applied to existing forms of human–machine interaction, especially those that occur in the context of the contemporary Internet and Web. Finally, I hope to demonstrate that the HEMC concept enables us to identify a number of important links between (currently) disparate areas of theoretical and empirical research.

2. Machine Cognition

One of the ways in which we might seek to the challenge the HEMC thesis is by questioning the foundational notion of *machine cognition*. We might thus seek to undermine the validity of the HEMC concept by questioning the extent to which machine-based processes could (ever?[4]) count as 'cognitive'. At the very least, we might wonder what are the conditions under which we are justified in seeing a specific instance of machine-based processing as a genuine instance of machine cognition. The challenge might be somewhat crudely characterized as follows:

> "This idea of cognitive processing being distributed across a motley collection of human and machine components is all very well, but show us what it is that makes such forms of information processing worthy of their 'cognitive' status. What is it that enables us to speak of some form of machine-based processing (extended or otherwise) as a genuine case of machine cognition?"

---

[4]For present purposes, I will assume, in accord with functionalist intuitions about the mind (Polger, 2009), that machine cognition is materially possible. In other words, I will assume that cognitive phenomena are, in principle at least, subject to non-biological realization.

This, it should be clear, is an important question. If we cannot identify a set of conditions that enable us to distinguish cognitive from non-cognitive phenomena, then the HEMC concept seems to be of limited scientific and philosophical value. Ideally, then, what is required is an understanding of what Adams and Aizawa (2010) dub the "mark of the cognitive" (i.e., the set of features that enable us to distinguish cognitive phenomena from their non-cognitive counterparts) (see also Adams, 2010; Adams & Garrison, 2013; Rowlands, 2009). Such an understanding looks to be of crucial importance when it comes to claims about the cognitive status of some episode of machine-based processing.

As is perhaps obvious from the fact that the issue is being raised, there is (as yet) no definitive answer to the question of what it is that makes a particular process a *cognitive* process (or, more generally, what makes some phenomenon a *cognitive* phenomenon). For their part, Adams and Aizawa attempt to limn the realm of the cognitive by appealing to the idea that "cognition is constituted by certain sorts of causal processes that involve nonderived content" (Adams & Aizawa, 2010, p. 67). Such a view leads Adams and Aizawa to countenance a form of contingent intracranialism about the (human) mind. They suggest that only neural processes, as a matter of contingent fact, are able to give rise to cognitive phenomena. Such claims, it should be clear, are largely inimical to ideas about extended cognition in the case of human cognizing, as well as (perhaps) the cognitive status of extant forms of machine intelligence. There are, however, a number of reasons to doubt the integrity of the claims made by Adams and Aizawa. Clark (2008; 2010a; 2010b), for example, has mounted a spirited defense of active externalism in the wake of Adams and Aizawa's critique. Such defenses, I suggest, are just as applicable to machine cognition (extended or otherwise) as they are to extended forms of human cognition.

Another view of cognition is proposed by Rowlands (2006). He suggests that a cognitive process should be seen as a process that is "(i) required for the accomplishing of a cognitive task, (ii) involves information processing, and (iii) is of the sort that is capable of yielding a cognitive state" (p. 32). One of the problems with this definition is that it introduces two new terms (i.e., "cognitive

task" and "cognitive state"), neither of which are particularly well understood. Rowlands suggests that we can resort to ostensive definitions in individuating cognitive tasks from their non-cognitive counterparts. However, inasmuch as we accept that cognitive tasks can be afforded ostensive definitions, then it is unclear why a similar ostensive strategy cannot be applied to the problem of discriminating between cognitive and non-cognitive processes.

Finally, Clark (2010a) has suggested that the notion of a cognitive process might be "best unpacked as the notion of a process that supports certain kinds of behavior (actual and counterfactual)" (p. 93). In this case, Clark appeals to the idea that our ability to individuate the realm of the cognitive is linked to the way in which a particular process contributes to the expression of intelligent behavior. He suggests that "[t]o identify cognitive processes as those processes, however many and varied, that support intelligent behavior may be the best we can do" (Clark, 2010a, p. 93).

I will not attempt to adjudicate between these different approaches to cognition in the present paper. What matters, for present purposes, is that none of these different ways of conceptualizing (human) cognition are incompatible

with the basic possibility of machine cognition.[5] Clark and Rowlands are both happy to place considerable emphasis on our intuitive grasp of what it is that makes something a cognitive process (e.g., a cognitive process is something that we identify as relevant to the expression of intelligent behavior), and Rowlands makes an additional claim about the importance of information processing to cognitive status. These claims, I suggest, are entirely compatible with the basic possibility of machine cognition. For all that the notion of machine cognition requires, in these cases, is (1) that we encounter a process that can be analyzed from a computational and information theoretic perspective, and (2) that we encounter a process that we are content to gloss as cognitive based on its contributions to intelligent behavior. Neither of these conditions, it should be clear, pose any problem for the claim that some kinds of machine-based system can function as cognitive systems. Indeed, the current interest in so-called *cognitive computing systems* (Hurwitz et al., 2015; Kelly & Hamm, 2013; Modha et al., 2011) is testament to the fact that at least some forms of intelligent system can be viewed in cognitive terms.

---

[5]One exception, perhaps, is the issue of non-derived content (or intrinsic intentionality), as discussed by Adams and Aizawa (2010). Adams and Aizawa may wish to reject the possibility of machine cognition on the grounds that the representational elements manipulated by a conventional computational system do not exhibit intrinsic intentionality (i.e., their contents are all derivative and fixed by convention). A couple of points are worth noting here. The first is that it is far from clear that intrinsic intentionality is beyond the reach of *all* forms of machine-based processing. Certain kinds of computational system (e.g., those whose computational economy is organized along more 'neural' lines) may be capable of trading in representations with non-derived content. Secondly, inasmuch as the proponents of intrinsic intentionality are willing to accept that only some of the representational elements of a cognitive system need to have non-derived contents, then HEMC may fit the bill even if non-extended forms of machine cognition do not. This is because, in the case of HEMC, we confront a system in which some of the representational elements of the larger system are likely to be located in the brains of human agents. Such representations, it seems, are ideally placed to exhibit intrinsic intentionality.

In order to help guide our intuitions about machine cognition, I suggest that it helps to draw on a philosophical instrument that was originally proposed by Clark and Chalmers (1998) as part of their seminal work on the active externalist position. This instrument comes in the form of what is called the *Parity Principle*:

Parity Principle

If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing [it] as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process. (Clark & Chalmers, 1998, p. 8; original emphasis)

The Parity Principle is intended to motivate the case for extended cognition by encouraging us to ignore issues of material constitution in determining whether or not some part of the physical world should be counted as part of the supervenience base for cognitively-relevant phenomena. We can use this approach, I suggest, to help guide our intuitions about the cognitive status of machine-based processing. I thus propose a variant of the Parity Principle that is intended to motivate the case for machine cognition:

Parity Principle for Machine Cognition (PPMC)

Whenever we encounter a putative case of machine cognition, we should ask ourselves whether the target process is one that we would be content to label as a cognitive process if such a process were—contrary to all the facts—to have occurred within the head of a human agent. If the answer to this question is "yes," then the process performed by the machine should be counted as a form of cognitive processing.

What this particular form of the parity principle asks us to do is imagine a state-of-affairs in which some episode of machine-based processing (e.g., machine vision) is subject to neurological realization. With this state-of-affairs in mind,

we can then ask ourselves whether we would be happy to view this counterfactual case of neural information processing as a genuine form of cognitive processing. If the answer to this question is "yes" (in the counterfactual case), then why should we revise our judgments about the cognitive status of the process just because it is realized (in the actual case) by some alternative nexus of material elements? As should be clear, the aim of the PPMC—just like the original Parity Principle—is intended to shield our evaluative efforts from any form of bio-centric prejudice or bias.[6]

3. Extended Cognition

The most popular expression of active externalist theorizing within the philosophy of mind comes in the form of *extended cognition* (or the *extended mind*) (Clark, 2008; Clark & Chalmers, 1998). The term "extended cognition" identifies an important and influential body of philosophical and cognitive scientific work that seeks to advance our understanding of the mechanistic underpinnings of the human mind. In particular, a key claim of the extended cognition movement is that the causally-active physical vehicles of human cognition can sometimes include (in addition to the standard neurological elements) a range of non-

---

[6]Just to be clear on this particular point, the issue is not whether a particular form of machine-based processing could be subject to neurological realization, i.e., implemented by the biological brain. Rather, the issue is whether we would regard such a process as cognitive if it were, in fact, implemented by the biological brain. It may, of course, be the case that certain forms of machine-based processing could never be subject to standard forms of neurological realization on the grounds of (e.g.) computational efficiency. In these cases, our intuitions as to the cognitive status of a process are perhaps, at best, unclear: we might conclude that the process is sufficiently 'alien' as to impugn it's cognitive status, or we might regard it as a novel form of cognitive processing that is worthy of further cognitive scientific attention. We should not thus assume that the PPMC gives us a "mark of the cognitive" for machine-based systems. Nor should we assume that the PPMC will be applicable to *all* forms of machine intelligence. In spite of these constraints, however, the PPMC may be of value in helping us determine the cognitive status of at least some forms of machine-based processing (see, for example, Section 7).

biological, extra-organismic resources, such as technological artifacts (Clark, 2008; Clark & Chalmers, 1998). Extended cognition is thus a claim about the nature of the realization base for some episode of intelligent behavior. In the non-extended case, the realizers of human mental states and processes are deemed to be the standard set of neurological realizers—the neural machinery of the intra-cranial realm. In the case of extended cognition, however, the realization base is deemed to be an extended one. In this case, mental states and processes are seen to supervene on a collection of materially-heterogeneous[7] elements, which includes both the usual biological suspects (i.e., the brain and body) as well as a set of non-biological resources (e.g., resources drawn from the surrounding social, physical and technological environment).

In order to help motivate the case for extended cognition, it helps to have a concrete example. With this mind, imagine that you are confronted with the task of explaining how a typical individual manages to perform the task of multiplying two four digit numbers. One option, here, is to focus on the role of neurally-encoded, imaginative manipulations of the problem-space. One kind of explanatory account might thus seek to explain how we first derive some symbolic encoding of the visual (or auditory) input corresponding to the two numbers. It would then appeal to a computational story, according to which the inner symbols are manipulated in such a way as to achieve the correct mathematical outcome.

Now contrast this sort of account with what is surely a more accurate (and ecologically-realistic) picture of how we implement long multiplication in the real world. This alternative picture involves the active manipulation of *external* symbols in such a way that the kind of problem confronting the biological brain

---

[7]In fact, claims of material heterogeneity are sometimes overplayed by proponents of active externalism. In the case of socially-extended cognition, for example, the cognitive processing routines of an individual human agent are materially extended as a result of their interactions with another *human* agent (Tollefsen, 2006). There is, in this case, no material heterogeneity in the underlying realizing mechanisms, since the mechanisms are all located in the neural (or at least the biological) realm.

is profoundly simplified. In place of the purely inner, environmentally-decoupled computational operations we now see a pattern of world-involving perception-action cycles, ones in which single digit numbers are compared and intermediate computational results are stored in an external medium using (e.g.) pen and paper.

This example, described in Wilson and Clark (2009), is a nice example of what might be called environmentally-extended computation or *wide computationalism* (see Kersten, 2017; 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, extra-organismic environment. A key feature of this world-involving explanatory account is that non-neural resources are sometimes seen to play a functional role that resembles that which would have been played by the neural elements if the process in question had been implemented solely in the inner, neural realm. In the case of long multiplication, for example, the paper can be seen to play the role of the visuo-spatial sketchpad component of working memory, while the human agent, in their interaction with the bio-external environment, can be seen to play a role akin to the central executive.

What the notion of extended cognition gives us, therefore, is the idea that extra-neural and extra-corporeal resources can, on occasion, form part of the explanatorily-relevant physical fabric that accounts for our success (or failure) in performing certain kinds of cognitive task. The explanatory story is, if you like, spread across the biological (brain and body) and non-biological (pen and paper) elements of an integrated nexus of information processing elements. This is important, for whenever we encounter a case of explanatory spread, we are also obliged to give explanatory weight to forces and factors that might otherwise have been deemed to be causally-irrelevant to some target phenomenon (see Wheeler & Clark, 2008). And in cases where the target phenomenon is a cognitive phenomenon (e.g., a cognitive process), explanatory spread will often require us to see the causally-relevant mechanisms as extending beyond the neural realm.

What seems to be important, then, is that in cases of extended cognition

we confront a set of components (brain, body and worldly elements) that are *connected* together in such a way as to ensure that they form part of a functionally-integrated (yet internally differentiated) system—one that is relevant to the mechanistic realization of some cognitive phenomenon of interest. This emphasis on connectivity establishes a potentially fruitful link with work in network science. In particular, it seems that at least some kinds of extended cognitive circuitry can be thought of as a *multi-modal* network[8] (Smart et al., 2010). We thus encounter a case of extended cognizing when the information processing circuits that realize a particular cognitive phenomenon are seen to extend beyond the traditional biological borders of the individual human agent. The time-variant structural profile of this 'extended' network is, of course, poised to influence the shape of specific (extended) cognitive performances, just as the time-variant structural profile of brain-based neural networks is poised to influence the shape of neurally-realized (non-extended) cognitive performances. In spite of all this, however, the biological agent is still seen to occupy a position of central importance when it comes to issues of cognitive ownership, control and responsibility (Clowes, 2015; Roberts, 2012). It is thus the (human) biological agent that is deemed to be responsible for the success (or failure) of a cognitive process, and it is the cognitive abilities of the (human) biological agent that are deemed to be enhanced as a result of the cognitive incorporation of extra-organismic elements. As we will see in Section 5, issues of ownership, control and responsibility are likely to be of similar importance when it comes to the attempt to evaluate putative cases of HEMC.

4. Human-Extended Machine Cognition

Now that we have a better understanding of what is meant by the terms "machine cognition" and "extended cognition," it is time to turn our attention

---

[8]A multi-modal network is a network that consists of multiple kinds of nodes, where the nodal elements are (typically) individuated with respect to material criteria (see Contractor et al., 2011).

to the concept of HEMC. Perhaps the best way of creeping up on this concept is to imagine a conventional case of extended cognition in which the cognitive performances of a specific human individual are materially-extended as a result of their interaction with some kind of computational device. Now imagine that instead of the device being factored into the processing routines of the human agent, it is, instead, the human agent that is factored into the processing routines of the computational device. In this case, it is the human agent that is supporting the device with respect to whatever tasks the device is attempting to perform. The concept of HEMC thus requires something of a 'mental flip' relative to conventional forms of extended cognition. In particular, HEMC requires us to reverse the usual perspective that we adopt with respect to the biological and technological elements of an extended cognitive organization. Instead of the non-biological elements being factored into the cognitive processing routines of a human agent, we are instead required to see the human agent as a resource that is incorporated into the cognitive processing routines of a synthetic (technological) system. A useful way of thinking about HEMC is thus to see it as a form of wide computationalism in reverse. In the case of HEMC, we are still confronted with a bio-technologically hybrid system that arguably participates in some form of cognitive processing. All that has changed, relative to the conventional case, is that our primary focus of attention has shifted away from the biological (i.e., human) elements and settled on the technological elements of the larger system. In other words, it is the 'cognitive' capabilities of the technological elements that we now see as being extended (and perhaps enhanced) as a result of the specific form of techno-biological bonding that takes place.

The HEMC concept, it should be clear, is a concept that is firmly rooted in contemporary philosophical debates about the extended character of human cognition. The HEMC thesis thus embraces the familiar notion of extended realization bases for processes that we recognize as, in some sense, cognitive in nature (see Section 2). What it adds to existing debates in this area is the idea of extended supervenience bases for the cognitive states and processes of machine-based systems. In other words, HEMC encourages us to take an

extended view of machine cognition. There are a number of reasons why this view is important.

Firstly, the HEMC concept is able to highlight the philosophical relevance of recent work regarding human computation systems (Law & von Ahn, 2011; Michelucci, 2013) and crowdsourcing techniques (Michelucci & Dickinson, 2016). A common focus of interest in these areas is the extent to which human agents can be *incorporated* into computational tasks that are managed, monitored and maintained by some form of technological (typically Web-based) system. A virtue of the HEMC concept is thus the way it helps to identify points of conceptual contact between otherwise disparate areas of research. Work on human computation, for example, has typically proceeded independently of work in philosophy of mind. Likewise work on extended cognition within the philosophy of mind and cognitive science has typically paid little attention to work on human computation. The HEMC concept provides us with a useful means of linking work within these areas, helping us detect points of interest that are shared by the computer science, cognitive science and philosophy of mind communities. One example concerns the value of bio-technological hybridization for cognitive and computational processing. Issues of complementarity and hybridization are a prominent focus of attention for proponents of extended cognition (Clark, 2006; Menary, 2010; Sutton, 2010), with the inner (i.e., neural) and outer (e.g., material artifact) resources often seen to make distinct, albeit complementary, contributions to cognitive processing. HEMC enables us to align this body of philosophical work with work in computer science, especially that relating to complementary computing (Kapoor et al., 2008), mixed-initiative systems (Horvitz, 2007), joint human–machine cognitive systems (Woods, 1985), interactive machine learning (Fogarty et al., 2008), symbiotic interaction (Jacucci et al., 2014), and heterotic computing (Kendon et al., 2015).

A second reason why the HEMC concept is important is because of its potential to broaden philosophical debates concerning extended cognition (and, more broadly, active externalism). A key example of this occurs in respect of what is called the Hypothesis of Organism-Centred Cognition (HOC) (Clark,

2008). This is the idea that:

> Human cognitive processing (sometimes) literally extends into the environment surrounding the organism. But the organism (and within the organism, the brain/CNS) remains the core and currently the most active element. Cognition is organism centered even when it is not organism bound. (Clark, 2008, p. 139)

The HOC thus situates a biological (human) organism at the heart of an extended cognitive system. This leads to a somewhat bio-centric and human-centric view of extended cognition. That is to say, the HOC encourages us to see extended cognition as something that is *always* centered on a particular biological agent, typically a human individual. In fact, this particular point serves as a source of controversy within the philosophy of mind (see Clark, 2011). Hutchins (2011), for example, argues that the HOC cedes too much to the biological brain, and that philosophers would do well to adopt a more decentralized and biologically-neutral perspective. But now notice that in the wake of claims about HEMC, we are able to identify a different sort of problem with the HOC. This problem is revealed once we recognize that, in cases of HEMC, it is no longer a human agent (or even a biological agent) that lies at the heart of an extended nexus of cognitively-relevant, information processing elements. Instead, the agent that lies at the heart of an episode of extended machine-based cognizing is, of course, a machine agent. Inasmuch as we accept the notion of HEMC, therefore, the HOC does, indeed, look to be overly bio-centric. A revised version of the HOC should probably attempt to respect the core idea of agential control, while simultaneously relinquishing any lingering vestige of bio-centric bias. In the wake of such revisionary efforts, which (importantly) are both instigated and informed by the concept of HEMC, the HOC may very well be transformed into something that is better able to ease the conceptual tensions that have arisen *within* the active externalist camp.

Finally, note that by extending the notion of extended cognition to the realm of machine cognition and AI systems, we are provided with a potentially

important opportunity to advance the state-of-the-art in machine intelligence. For if it is indeed our ability to engage in cognitively-potent bio-technological unions that best explains human-level cognitive success (see Clark, 2003), then perhaps the route to synthetic forms of human-level intelligence lies in the development of computational systems that are able to press maximal cognitive benefit from the environments in which they are materially embedded.

5. When is Machine Cognition Extended?

Assuming that we accept the basic possibility of HEMC, we need to address an important question: When is a particular instance of machine-based processing a *bona fide* case of HEMC? In other words, what are the conditions under which we should see an episode of machine-based processing as materially-extended?

In addressing this issue, I suggest it pays to consider the following criteria. We encounter a case of HEMC, I suggest, when:

1. we are confronted with a process that we are prepared to accept as a cognitive process,

2. the process in question is realized by a mixture of human and technological elements (i.e., the realization base for the process is bio-technologically hybrid), and

3. the technological elements serve as a particular focus of attention when it comes to issues of agential control and coordination (this helps to ensure that it is the cognitive profile of the technological system that is being enhanced by the human elements rather than the other way round).

Unfortunately, none of these criteria are entirely unproblematic. The first criterion obviously raises issues about the cognitive status of tasks and processes, which, as we have seen, are a point of contention in the philosophical community (see Adams & Aizawa, 2010; Clark, 2010a). Given that we have already discussed this issue at some length (see Section 2), let us direct our attention to the second criterion—the criterion that deals with extended realization bases. This is the (by

now familiar) idea that the vehicles of some cognitive process can extend beyond the typical (biological or technological) borders of a cognitive agent. In the case of HEMC, the idea translates to the claim that human agents should be seen as part of the supervenience base for machine-based cognitive states and processes. In order to help us gain a better understanding of this criterion, imagine that you encounter a seemingly self-contained cognitive computing system that exhibits an intelligent processing capability. Presumably, you would, in this case, be content to see the material elements of the system (the circuit boards, central processing units, storage disks, and so on) as part of the physical substrate (the physical machinery) that realizes the 'cognitive' capabilities of the target system. But now imagine that a particular part of the computational process—a particular subroutine, let's say—is performed by a human agent instead of the usual medley of technological elements. The involvement of the human agent, in this case, let us assume, makes no difference to the actual performance profile of the larger system: the human agent thus takes the same amount of time to perform the task and produces outputs that are identical to those that would have been observed in the purely technological case. With this state-of-affairs in mind, we can now ask ourselves whether the human agent should be included in the network of material elements that *realizes* the performances of the target cognitive computing system. I suggest that in such situations we should, indeed, see the human agent as part of the physical fabric that realizes the performances in question. The reason for this, I suggest, relates to the functional role of the human agent in the larger cognitive economy of the system. Given that the functional contribution of the human agent is the same as that which would have been performed by a technological element (in the non-extended case), it seems fair to conclude that the human agent is (from a functional perspective, at least) no different to any other component of the relevant system. Functional parity is thus one of the things that guides our sense as to whether the human agent is part of the realization base for some episode of machine-based cognizing. If the cognitive computing system had resorted to a back-up processor as a means of performing the desired computation (the one undertaken by the human agent),

we would, I assume, have had no problem in seeing the back-up processor as part of the physical fabric that realized the target process of interest. But given that the human agent is playing a role that is functionally identical to that performed by the backup processor, why should our view of the human agent, in the hybrid case, be any different? If we are content to see the backup processor as part of the realization base for a particular performance, then why not see the human agent as part of the realization base for what is, in effect, the same performance?

If this example doesn't grab you, perhaps it will help to consider a more concrete case. Let us therefore turn our attention to an existing system called Galaxy Zoo (Lintott et al., 2008). Galaxy Zoo is a Web-based citizen science system that aims to support the classification of a large number of galaxies, as imaged by the Sloan Digital Sky Survey. As it is described by Lintott et al. (2008), the Galaxy Zoo system relies on the ability of human agents to classify galaxies based on a range of morphological criteria. But now let us suppose that in the wake of recent developments in computer vision technology (see Dieleman et al., 2015), the services of the human agents are no longer required. In this case, *all* the relevant classification processes are ported to the technological realm and are thus subject to conventional forms of computational realization. Now, inasmuch as we accept that the galaxy classification task is a cognitive task, there seems little reason to reject the claim that the automated (computerized) version of the task is also a cognitive task—it is, after all, the same basic task. In addition, there can be little doubt that, in this situation, we should view the technological elements of the machine-based system as relevant to the mechanistic realization of the system's classificatory capabilities (for what else could be realizing the process in question?). But now note that once we concede these points, it seems reasonable to conclude that the original bio-technologically hybrid system was, in fact, implementing a cognitive process that was realized by a combination of human and machine elements.

Now let us turn our attention to the third of the aforementioned HEMC criteria—the one relating to agential control. This criterion is perhaps the most problematic for the HEMC thesis. The general idea is that we should

see some set of technological or machine-based elements as being credited with the ownership or possession of specific cognitive capabilities. This is where issues of extended cognition come face-to-face with issues of cognitive agency. In the case of human cognition, it is the human individual (the human agent) that is the focus of our ascriptive endeavors regarding the possession of specific cognitive capabilities. This is so, even in situations where the cognitive states and processes of the relevant individual are subject to material extension. Thus, even in cases of extended cognition, it is still the biological person that deserves credit for achieving a specific cognitive outcome (e.g., writing an academic paper), and this is despite the fact that such achievements are often seen to rely on an extended nexus of extra-organismic (i.e., extra-agential) elements (see Clark, 1997, pp. 206–207). Similarly, when it comes to ascriptions of knowledge, it is 'I'—the biological agent—that deserves credit (or blame) for the beliefs that I hold, despite the fact that the physical machinery that underlies my epistemic capabilities may, on occasion, include all manner of non-biological props, aids and artifacts (see Smart, in press). Our ascriptions of skill, knowledge and cognitive ability thus seem to pick out a particular part of the causally-relevant physical matrix that realizes some phenomenon of interest. Such ascriptions, I suggest, are guided by our sense of what parts of the system are responsible for influencing the flow of information within the system, as well as (perhaps) the parts of the system that are doing the bulk of the information processing work.

There is, therefore, an appeal to issues of cognitive agency when it comes to our understanding of HEMC. In the human case, we see the biological agent as playing a crucial role in creating, coordinating and controlling the circuits that manage to bring an extended cognitive organization into being. As Clark (2008) notes:

> Just as it is the spider body that spins and maintains the web that then (following Dawkins 1982) constitutes part of its own extended phenotype, so it is the biological human organism that spins, selects, or maintains the webs of cognitive scaffolding that participate in the

20

extended machinery of its own thought and reason. (p. 123)

In the HEMC case, of course, we can reverse the usual order of things and see the machine elements as playing a role akin to that played by the human agent in more conventional (i.e., human-centered) cases of cognitive extension. In particular, when it comes to HEMC, it is the machine-based system that plays a substantive role in creating, coordinating and controlling the extended circuitry that realizes cognitively-relevant performances. As is noted by Clark (2015; 2016), a useful way of thinking about these processes of creation, coordination and control is to perhaps see them as *meta-cognitive* processes, i.e., as processes that embody knowledge concerning, among other things, the relationship between different information processing strategies and their associated outcomes. The function of these meta-cognitive processes (in both the human and machine cases) is thus to monitor and control the shape of cognitive processes in ways that meliorate the chances of cognitive success. The processes are, in effect, supporting the selection and adoption of particular cognitive strategies based on the nature of the task at hand and the relative strengths and weaknesses of whatever ecological resources (biological and non-biological) are available for use.

When it comes to HEMC, therefore, it is important that machine-based elements are seen to bear the bulk of the responsibility for the creation, monitoring and maintenance of the information processing loops that incorporate one or more human agents into the machine's own cognitive processing routines. This is important, I suggest, because we need to be sure that it is the capabilities of the machine-based system that are materially-extended (and hopefully enhanced) rather than the other way around (i.e., the technological elements enhancing the cognitive capabilities of the human agent). A HEMC system is thus a system in which the machine elements are the primary focus of our attention when it comes to issues of agential control and coordination. Such ascriptions of ownership, responsibility and control may very well be based on the extent to which we see some subset of elements as playing an *active* role in shaping the information

processing profile of the larger systemic organization. We will encounter some specific examples of these forms of machine-based 'meta-cognitive control' in later sections (see Section 7 and Section 8).

6. Neural Wideware

The term "wideware" is used by proponents of extended cognition to refer to the physical elements that comprise an extended cognitive circuit (Clark, 1999). Typically, of course, it is assumed that the elements targeted by the notion of wideware are of a non-neural nature. This is because the conventional focus of active externalist theorizing concerns the elements that lie outwith the neural realm. In discussing the notion of wideware, for example, Clark (1999) suggests that we should see the term as referring to an item that:

> ...must be in some intuitive sense environmental: it must not, at any rate, be realized within the biological brain or the central nervous system. Bodily aspects and motions, as well as truly external items such as notebooks and calculators, thus fit the bill. (p. 11)

But now notice how a consideration of HEMC challenges this conventional portrayal of the wideware concept. In the case of HEMC, it no longer makes any sense to exclude neural elements from the suite of resources that can participate in the realization of extended cognitive processing routines. This is because, in the case of HEMC, it seems that the human biological brain could itself form part of the supervenience base that realizes the cognitive states and processes of a machine-based system, such as a cognitive computing system, a robotic platform or some other form of AI system. A consideration of HEMC thus encourages us to consider the possibility of *neural wideware*: a specific form of wideware in which it is the biological brain of a human individual that is playing an explanatorily-significant role in accounting for the cognitive capabilities ascribed to another entity (either another human agent or an AI system).

In order to help us develop a better understanding what is being proposed here, consider Clickworkers (Kanefsky et al., 2001), one of the early demonstrations of

the power and potential of crowdsourcing techniques for the purposes of citizen science. Clickworkers was a project that sought to engage human volunteers in the task of identifying craters on the surface of Mars, based on images taken by the Mars Global Surveyor spacecraft. As part of the task, users were presented with a series of satellite images and asked to manually annotate the images so as to record crater locations.

As it is presented, the Clickworkers crater mapping task is clearly one that relies on the visuo-cognitive capabilities of human subjects; in particular, their ability to identify distinct visual features (i.e., craters) from a set of satellite images. There should, in this case, be little doubt as to the importance and relevance of the humans' biological brains in generating the informational outcomes that are delivered back to the machine-based system for further processing. But now consider an alternative approach to the realization of the crater mapping task. This approach would, let us assume, involve a form of image preprocessing in which the raw photographic images were first modified to highlight the relevant features of interest (e.g., the crater outlines). The images would then be subject to automated image segmentation routines that decomposed the original image array into a set of image tiles.[9] Next let us imagine that the processed image tiles are presented to a group of human volunteers using the rapid serial visual presentation paradigm described by Huang et al. (2011). Here, a series of images are presented in quick succession (e.g., at a rate of 5 images per second) to a human subject, and EEG recording techniques are used to detect the event-related potentials that indicate the presence of target features. Such techniques draw on the pattern recognition capabilities of the biological brain in order to process a large number of images within a relatively short timeframe (e.g., at a rate of 5 images per second an individual can process 300 images per minute). The final step in the processing routine is to feed the results of the EEG recordings back to the machine-based system, whereupon

---

[9]Both of these processes are, in fact, demonstrated by recent attempts to apply image processing techniques to the crater mapping task (see Barata et al., 2004).

the machine attempts to reconcile the recordings with the results obtained from other human subjects, as well the outputs of its own image processing algorithms (see Barata et al., 2004).[10]

The resulting (hypothetical) system is, I suggest, a system that exhibits a form of HEMC. The system is clearly one in which we encounter a productive interplay between the processing capabilities of material elements that are spread across both the biological and technological domains. It is, moreover, a system that taps directly into the neuro-computational capacities of the human biological brain. In this case, it seems appropriate to say that the cognitive routines of the machine-based system are ones that supervene, at least in part, on the biological brains of the human subjects. As a means of reinforcing this particular point, suppose that the state-of-the-art in image processing advanced to the point where the kinds of operations performed by the human participants of the aforementioned system were able to be implemented by a suite of fully automated image analysis and annotation services. Would we, in this situation, be inclined to say that the image analysis services should *not* be seen as a legitimate part of the physical machinery that realized the crater mapping task? This seems unlikely. But once we accept this point, then there seems little reason to reject the claim that the brains of the human agents in the bio-technologically hybrid case are functioning as part of the material fabric that realizes the relevant process.

HEMC thus opens the door to claims about neural wideware—the idea that neurological elements can, on occasion, serve as constituent elements of an extended cognitive circuit.[11] It might be thought that this idea is one that is in

---

[10]Although this is a purely hypothetical system, the capabilities being described here are largely consistent with a body of work that seeks to make use of the human brain as a means of enhancing the current state-of-the-art in image processing (see Mohedano et al., 2014; Wang et al., 2015).

[11]Such claims, it seems to me, make perfect sense in the wake of research concerning so-called neurocomputers. Here, the computational capabilities of some neural resource (typically an *in vitro* preparation) are used to assist with the information processing tasks performed by

some tension with the original, Clarkian notion of wideware presented above—Clark (1999), recall, appears to suggest that the notion of wideware excludes the possibility of brain-based forms of realization. The tension, however, is merely illusory; for in developing the wideware concept, Clark's (1999) focus of interest lies squarely in the realms of individual human cognition. It is, as a result, somewhat unfair to see claims of HEMC—and thus of neural wideware—as representing a form of 'philosophical assault' on the original wideware concept. A more productive approach, in my view, is to see the present discussion as expanding our understanding of the various forms of wideware that can exist in a variety of different contexts. The notion of neural wideware thus helps us see how the original wideware concept can be adapted and applied to a 'new' kind of cognitive organization, one that perhaps lies beyond the typical purview of the basic active externalist position.[12]

7. Humans in the Loop

One response to the HEMC concept is to see it as identifying a futuristic class of AI systems that, quite literally, have people as their parts. This is certainly one way to view the HEMC concept. The value of the HEMC concept is, however, also apparent in the way it helps us approach a number of *existing* systems. Consider, for example, a system described by Branson et al. (2014; 2010), which relies on a combination of machine vision algorithms and human
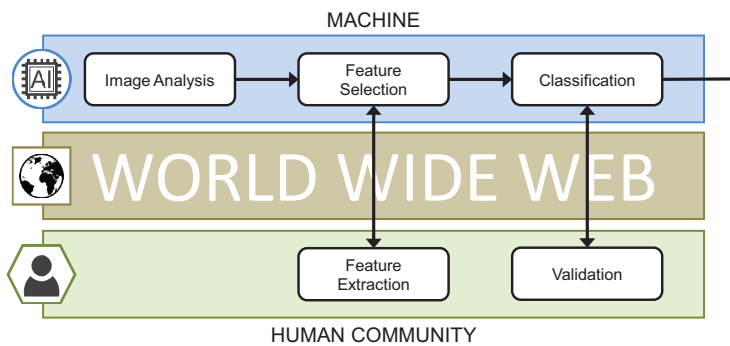
---

some larger system (e.g., a robotic agent) (DeMarse & Dockendorf, 2005; DeMarse et al., 2001; Pizzi et al., 2009; Ruaro et al., 2005).

[12]Interestingly, the notion of neural wideware looks to be applicable to debates regarding socially-extended cognition. Socially-extended cognition is a particular form of cognitive extension in which one or more human individuals are deemed to form part of the extended cognitive circuitry of another human individual (e.g., Tollefsen, 2006). HEMC, in this case, emerges as a particular form of socially-extended cognition. The only difference with conventional (i.e., human-centered) forms of socially-extended cognition is that HEMC replaces the central human agent (i.e., the human agent whose cognitive capabilities are deemed to be extended) with a cognitive agent of the technological (or at least non-human) variety.

Finch? Bunting? Blue Belly? Conical Beak? If {feature} Then {Category}

Hard for humans   Easy for humans   Hard for humans
Hard for machines   Hard for machines   Easy for machines

(a)

MACHINE

AI → Image Analysis → Feature Selection → Classification →

WORLD WIDE WEB

Feature Extraction   Validation

HUMAN COMMUNITY

(b)

Figure 1: The classification of images based on the species of bird shown in the image is a task that is difficult for both humans and machines (a). The larger task can, however, be broken down into a series of smaller steps and assigned to the human and machine elements of a functionally-integrated socio-technical system (b). The result is a hybrid knowledge-based system that relies on the distinctive (and, in this case, complementary) capabilities of the human and machine elements.

input to perform a complex image classification task. Branson et al. were interested in developing a system that could identify the species of bird depicted in a series of photographic images. This, it should be clear, is a task that is difficult for both humans and machines to perform. When presented with the photograph shown in Figure 1a, for example, it is unlikely that many people (excluding those with ornithological expertise) would be able to identify the bird depicted in the photograph as the Indigo Bunting (*Passerina cyanea*). Similarly, the task of species identification is one that challenges the capabilities of contemporary machine vision systems. The key insight of Branson et al. was to recognize that the task of species identification could be decomposed into a

series of smaller, more tractable steps, each of which could be assigned to either the human or machine elements of a larger bio-technologically hybrid system (see Figure 1b). For example, one of the steps in the species identification task relates to the extraction of specific features. These include the color of the bird's plumage ("Does the bird have a blue belly?") and the shape of the bird's beak ("Does the bird have a beak that is conical in shape?"). Extracting these features from a natural scene using automated techniques is a task that is notoriously difficult for machine-based systems; however, it is a task that is relatively easy for humans to perform. The result is that the feature extraction task is one that ends up being delegated to a community of human agents. The delegation task, however, is one that is itself far from straightforward. In particular, the efficiency of the classification process is linked to the optimal selection of specific feature-related questions. There is, for example, no point in attempting to solicit information about beak shape if the machine-based system can already infer (on the basis of previous processing) that the beak can only be of one particular shape. Similarly, there is no point in soliciting input about features that do not reduce the probability distribution associated with candidate classifications. The selection of appropriate feature-oriented questions is thus something of a knowledge-intensive task in its own right—one that requires an ability to calculate the relative optimality (in an information theoretic sense) of different feature-related question sequences. Conventional computers, of course, are able to deal with the mathematical challenges posed by this particular problem, and it is for this reason that the task (or sub-task) of feature selection is one that ends up being assigned to the machine-based elements of the larger system (see Figure 1b).

The upshot of all this is that we have a hybrid intelligent system that interleaves the activity of multiple human agents with a collection of machine-based processing routines. Such a system is, I suggest, a system that trades in a genuine form of HEMC. Firstly, there can be little doubt that the target task is one that we should recognize as cognitive in nature. A human agent who performed the task in the absence of any kind of technological assistance

would undoubtedly be seen to be engaged in a cognitive task. In addition, the successful performance of the task would invariably be deemed to indicate the presence of some sort of cognitive ability. The fact that the task is, in the current case, realized by a nexus of materially-heterogeneous elements should not make any difference to our judgments regarding the actual nature of the task that is being performed. Here we see the value of the Parity Principle for Machine Cognition (see Section 2) in guiding our intuitions as to the cognitive status of a particular task.

Secondly, there can be little doubt that the realization base for the cognitive performances are, in the majority of cases,[13] bio-technologically extended. If we were to replace the contributions of the human agents with an intelligent (albeit automated) Web service that responded in *exactly* the same manner as the human agents, we would have no problem, I suggest, in concluding that the Web service should be seen as part of the physical fabric that realizes the larger task of species identification.

Finally, the technological elements of the system are playing an *active* role in shaping the organization of the information processing loops that determine the performance profile of the larger system. In particular, it is the technological elements that determine *what* sort of questions get presented to the human community, and it is the technological elements that determine *when* these particular questions are actually asked. These questions are ones that not only respect the specific capabilities (and limitations) of the human agent, they are also ones that work in concert with the machine's own information processing capabilities so as to solve the target problem in the most efficient manner. Crucially, the system designed by Branson et al. exhibits a form of what we might call *active engagement*. By this I mean that the nature of the informational contact that is established with the human user community

---

[13]In some cases, the task was solved *solely* by the machine-based components of the system. This occurred in cases where the machine vision algorithms were able to process the target image and yield an acceptable result.

is determined dynamically (at runtime) as a specific result of the machine's evaluation of the current problem state. It is thus the machine that is responsible for selecting questions that are likely to be the most informative based on the results of previous information processing. It is also the machine that is primarily responsible for determining the structure of the processing loops that realize the target task. In simple cases—where the machine vision algorithms are able to return an answer with high confidence—no human agents are recruited into the task. In more difficult cases, however, the machine might establish connections with the human community as a means of improving its confidence in particular response options. And in the most difficult cases, the machine vision algorithms might be of no use whatsoever in terms of identifying the relevant bird species. In these cases, the machine agent is utterly dependent on the responses of the human participants to derive a correct solution to the image processing problem. But even in these situations, the machine is far from cognitively inert with regard to the performance of the task. In particular, the machine is still required to select specific questions and process the responses it receives from the human agent community.

Here, then, we can see that the technological components of Branson et al.'s system are playing a crucial role in controlling the organization of the information processing circuits that realize the cognitive performances of the larger systemic organization. In particular, the technological components are playing a crucial role in determining what is sometimes referred to as the "effective connectivity" of the system (see Friston, 2011). This, I suggest, suffices for us to see the technological components as the primary focus of attention when it comes to issues of agential control and coordination. It is, I suggest, the technological components that determine the moment-to-moment configuration (and, indeed, the moment-to-moment material constitution) of the information processing system that realizes the target task, and it is, as a result, the technological components that should bear the *primary* credit (or blame) for whatever task-related successes (and failures) are exhibited by the larger, bio-technologically hybrid ensemble.

The result of all this is that *all* of the HEMC-related criteria presented in Section 5 are now satisfied: we have a discernible cognitive task, the task features the involvement of both human and machine elements, and the primary locus of agential control seems to reside with the machine elements as opposed to the human elements. In view of all this, I suggest that we should see the system described by Branson et al. (2014) as providing us with a real-world example of HEMC.

8. Ecological Assembly

One of the interesting features of Branson et al.'s system is that it exhibits of form of contingent connectivity, i.e., connections with the human community are only established in situations where the machine's own visual processing routines are deemed to be inadequate (something that is determined by the machine itself). In this sense, Branson et al.'s system can be seen to participate in processes that are relevant to the *creation* of extended cognitive circuits. In particular, when the system detects a state-of-affairs in which its own capabilities are likely to be deficient, it automatically switches to a mode of operation that factors in the contributions of the human community.

All of this will no doubt strike a chord with those who work in the area of extended cognitive systems. For issues concerning the creation or assembly of extended cognitive circuits are a popular focus of attention for both the supporters and detractors of active externalism. Often, issues of creation and assembly arise in relation to the HOC (see Section 4); however, they also surface in relation to what is called the Principle of Ecological Assembly (PEA) (Clark, 2008). The basic idea behind the PEA is that human cognizers are seen to engage in the active selection (and incorporation) of resources that best meet their requirements in respect of a particular task. According to the PEA, "the canny cognizer tends to recruit, on the spot, whatever mix of problem-solving resources will yield an acceptable result with a minimum of effort" (Clark, 2008, p. 13). The image that emerges, here, is one in which the human biological agent

is seen to play an active role in the creation of the world-involving informational circuits that subsequently realize some aspect of the agent's cognitive processing.

It is here, I suggest, that we begin to see the philosophical and cognitive scientific relevance of a number of strands of work that are the current focus of attention in contemporary computer science. For the HEMC-based variant of the PEA encourages us to see the process of assembly as one that is undertaken by machine-based systems, specifically for the purposes of improving or enhancing *their* performance profile. The assembly process is also one that is oriented to the human social environment. In other words, when it comes to HEMC, the ecological backdrop to the assembly process is a world populated by the likes of you and me—we humans are thus the resources that are, in one way or another, co-opted into the information processing routines that work to extend the cognitive reach of some of our most advanced forms of technological contrivance.

Such insights dovetail with the focus of recent work in computer science, especially that which goes under the heading of human computation (Law & von Ahn, 2011; Michelucci, 2013). This includes work that aims to improve the access that machine-based systems have to the human social environment, typically via the technological infrastructure of the Web and Internet. Consider, for example, recent work relating to what is called the "human cloud" (Kaganer et al., 2013; Miller et al., 2010). The human cloud, in this case, is the human analogue of the conventional arsenal of online computing services that aims to provide individuals and organizations with remotely-located information processing and storage capabilities (Hayes, 2008). In essence, the notion of the human cloud introduces us to the idea that the human social environment can be seen as something of a computational resource in its own right, one that can be used to assist with certain kinds of information processing activity and (perhaps) the storage of certain kinds of information. This is useful, not just because it helps us think of the human social environment as a potential target of techno-biological bonding, but also because it enables us to appreciate the philosophical relevance of an array of current engineering efforts, especially those

that seek to harness the representational and computational potential of the global human community. Here, we see the use of service-oriented techniques as a means of supporting various forms of socio-computational processing (Schall, 2013; Schall et al., 2008), the extension of traditional Web service description languages to accommodate the possibility of human involvement (Schall, 2013), and the emergence of programming frameworks that are specifically intended to deliver "complex computation systems incorporating large crowds of networked humans and machines" (Minder & Bernstein, 2012, p. 124). Progress in these various areas of research is likely to improve the extent to which machine-based systems can select and recruit human agents as part of their attempt to assemble extended cognitive circuits. The value of the PEA, in this respect, is that it helps us see how a range of contemporary research and development efforts are providing a technological substrate that is largely conducive to the emergence of future forms of HEMC.

9. Conclusion

HEMC is a specific form of extended cognition that is applicable to the realm of AI. In particular, the HEMC thesis states that machine-based cognitive states and processes may, on occasion, be realized by a nexus of material elements that includes one or more human agents. Cast in this light, the HEMC concept enables us to extend the reach of conventional active externalist theorizing to the realm of AI systems and machine intelligence. Beyond this, however, the concept of HEMC enables us to entertain the possibility that *human agents* could form part of the physical machinery that realizes the cognitive capabilities of certain kinds of machine. The idea, in essence, is that certain forms of machine intelligence could be grounded in a mechanistic substrate that includes human agents as constituent elements. This is an idea that dovetails with the focus of recent work in computer science, especially that concerning hybrid systems (Demartini, 2015), human computation (Law & von Ahn, 2011; Michelucci, 2013), heterotic computing (Kendon et al., 2015), crowd computing (Murray et al.,

32

2010), and collective intelligence (Bigham et al., 2015; Weld et al., 2015).

In the present paper, I have attempted to provide an initial defense of the HEMC thesis, arguing that we should accept the idea that human agents can, on occasion, form part of the physical machinery that realizes the cognitive states and processes of machine-based systems. I have, moreover, sought to identify the conditions under which should recognize the existence of HEMC (see Section 5), and I have subsequently applied these conditions to evaluate a specific real-world system (see Section 7). As a result of this analysis, I suggested that the HEMC concept is more than just a matter of theoretical interest and speculation for the philosophy community, it is also concept that is potentially applicable to an ever-expanding array of socio-computational and socio-technical systems.

The HEMC concept has implications that are spread across a number of disciplines. Firstly, the concept is of interest and value to the philosophy of mind community, especially for those who embrace active externalist approaches to the mind. Here, the HEMC thesis promises to progress (or at least stimulate) philosophical debates regarding the notion of extended cognition. We saw, for example, that the HEMC thesis poses a challenge to the HOC (see Section 4), and we saw how it expands our understanding of the kinds of resources that can be incorporated into extended cognitive routines (see Section 6). The HEMC thesis also supports the effort to develop a conceptual taxonomy of extended cognitive systems (see Wilson & Clark, 2009). Thus, in addition to conventional forms of extended cognition (i.e., those centered on a single human individual), we are now in a position to consider a separate class of systems, *viz.*, those centered on some form of technological agent. Finally, the HEMC concept enables us to revisit (and perhaps reinvigorate) philosophical debates regarding the notion of socially-extended cognition; for cognitive extension of the HEMC variety is easily construed as a specific form of socially-extended cognition (see Section 6).[14]

---

[14]This is important because the Internet and Web might be seen to support the practical realization of HEMC (see Section 7) in a manner that is seldom, if ever, the case with

The HEMC concept is also important in the sense that it helps to reveal points of contact between otherwise disparate areas of research. With the concept of HEMC to hand, we are thus able to appreciate the philosophical significance and relevance of work relating to (e.g.) human computation (Law & von Ahn, 2011; Michelucci, 2013) and the human cloud (Miller et al., 2010). We are also able to gain a better understanding of the cognitive significance of the Internet and Web. For example, the cognitive significance of the Internet has typically been judged relative to its actual or potential effects on human cognition (Carr, 2010; Smart et al., 2017; Staley, 2014). However, when we look at the Internet through the conceptual lens of HEMC, we are afforded a rather different (and novel) perspective. In particular, we can see the Internet as providing an unprecedented form of access to the human social environment, profoundly altering the opportunities that are available for cognitively-potent forms of techno-biological bonding. Inasmuch as we link human intelligence to an ability to exploit the features of the environment in which humans are materially embedded (Clark, 2003), then perhaps the advent of the Internet should be seen as marking an important milestone in the development of AI systems. For if it is indeed the case that intelligence is as much about the environment in which an agent is embedded as it is the properties of the agent itself (see Simon, 1996), then perhaps there is no better way to advance the cause of AI than to embed our future machines in an environment that provides ever-more intimate forms of cognitive contact with a significant proportion of humanity.

Acknowledgements

---

socially-extended cognition of the human-centered kind (see Clark & Chalmers, 1998).

References

Adams, F. (2010). Why we still need a mark of the cognitive. *Cognitive Systems Research*, *11*, 324–331.

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

Adams, F., & Garrison, R. (2013). The mark of the cognitive. *Minds and Machines*, *23*, 339–352.

Barata, T., Alves, E. I., Saraiva, J., & Pina, P. (2004). Automatic recognition of impact craters on the surface of Mars. In *International Conference of Image Analysis and Recognition*. Porto, Portugal.

Bigham, J. P., Bernstein, M. S., & Adar, E. (2015). Human–computer interaction and collective intelligence. In T. W. Malone, & M. S. Bernstein (Eds.), *Handbook of Collective Intelligence*. Cambridge, Massachusetts, USA: MIT Press.

Branson, S., Horn, G., Wah, C., Perona, P., & Belongie, S. (2014). The ignorant led by the blind: A hybrid human–machine vision system for fine-grained categorization. *International Journal of Computer Vision*, *108*, 3–29.

Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P., & Belongie, S. (2010). Visual recognition with humans in the loop. In *11th European Conference on Computer Vision*. Heraklion, Crete, Greece.

Carr, N. (2010). *The Shallows: How the Internet Is Changing the Way We Think, Read and Remember*. London, UK: Atlantic Books.

Chrisley, R. (2003). Embodied artificial intelligence. *Artificial Intelligence*, *149*, 131–150.

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

Clark, A. (1999). Where brain, body, and world collide. *Cognitive Systems Research*, *1*, 5–17.

Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence*. Oxford, UK: Oxford University Press.

Clark, A. (2006). Material symbols. *Philosophical Psychology*, *19*, 291–307.

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

Clark, A. (2010a). Coupling, constitution and the cognitive kind: A reply to Adams and Aizawa. In R. Menary (Ed.), *The Extended Mind*. Cambridge, Massachusetts, USA: MIT Press.

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

Clark, A. (2011). Finding the mind. *Philosophical Studies*, *152*, 447–461.

Clark, A. (2015). What 'extended me' knows. *Synthese*, *192*, 3757–3775.

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action and the Embodied Mind*. New York, New York, USA: Oxford University Press.

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

Clowes, R. W. (2015). Thinking in the cloud: The cognitive incorporation of cloud-based technology. *Philosophy & Technology*, *28*, 261–296.

Contractor, N., Monge, P., & Leonardi, P. M. (2011). Multidimensional networks and the dynamics of sociomateriality: Bringing technology inside the network. *International Journal of Communication*, *5*, 682–720.

Craver, C., & Tabery, J. (2016). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Stanford, California, USA: Stanford University. (Spring 2016 ed.).

DeMarse, T. B., & Dockendorf, K. P. (2005). Adaptive flight control with living neuronal networks on microelectrode arrays. In *IEEE International Joint Conference on Neural Networks*. Montreal, Quebec, Canada.

DeMarse, T. B., Wagenaar, D. A., Blau, A. W., & Potter, S. M. (2001). The neurally controlled animat: Biological brains acting with simulated bodies. *Autonomous Robots*, *11*, 305–310.

Demartini, G. (2015). Hybrid human–machine information systems: Challenges and opportunities. *Computer Networks*, *90*, 5–13.

Dieleman, S., Willett, K. W., & Dambre, J. (2015). Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly Notices of the Royal Astronomical Society*, *450*, 1441–1459.

Fogarty, J., Tan, D., Kapoor, A., & Winder, S. (2008). CueFlik: Interactive concept learning in image search. In *Conference on Human Factors in Computing Systems*. Florence, Italy.

Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain Connectivity*, *1*, 13–36.

Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, *173*, 466–500.

Harnad, S., & Dror, I. E. (2006). Distributed cognition: Cognizing, autonomy and the Turing test. *Pragmatics & Cognition*, *14*, 209–213.

Hayes, B. (2008). Cloud computing. *Communications of the ACM*, *51*, 9–10.

Horvitz, E. J. (2007). Reflections on challenges and promises of mixed-initiative interaction. *AI Magazine*, *28*, 19–22.

Huang, Y., Erdogmus, D., Pavel, M., Mathan, S., & Hild, K. E. (2011). A framework for rapid visual image search using single-trial brain evoked responses. *Neurocomputing*, *74*, 2041–2051.

Hurwitz, J. S., Kaufman, M., & Bowles, A. (2015). *Cognitive Computing and Big Data Analytics*. Indianapolis, Indiana, USA: John Wiley & Sons, Inc.

Hutchins, E. (2011). Enculturating the supersized mind. *Philosophical Studies*, *152*, 437–446.

Iida, F., Pfeifer, R., Steels, L., & Kuniyoshi, Y. (Eds.) (2004). *Embodied Artificial Intelligence*. Berlin, Germany: Springer-Verlag.

Jacucci, G., Gamberini, L., Freeman, J., & Spagnolli, A. (Eds.) (2014). *Symbiotic Interaction*. London, UK: Springer.

Jonker, C. (2008). Extended mind from an artificial intelligence perspective. In L. Zonneveld, H. Dijstelbloem, & D. Ringoir (Eds.), *Reshaping the Human Condition: Exploring Human Enhancement*. The Hague, Netherlands: Rathenau Institute.

Kaganer, E., Carmel, E., Hirschheim, R., & Olsen, T. (2013). Managing the human cloud. *MIT Sloan Management Review*, *54*, 23–32.

Kanefsky, B., Barlow, N. G., & Gulick, V. C. (2001). Can distributed volunteers accomplish massive data analysis tasks. In *32nd Lunar and Planetary Science Conference*. Houston, Texas, USA.

Kapoor, A., Tan, D., Shenoy, P., & Horvitz, E. (2008). Complementary computing for visual tasks: Meshing computer vision with human visual processing. In *8th IEEE International Conference on Automatic Face & Gesture Recognition*. Amsterdam, The Netherlands.

Kelly, J. E., & Hamm, S. (2013). *Smart Machines: IBM's Watson and the Era of Cognitive Computing*. Chichester, West Sussex, UK: Columbia University Press.

Kendon, V., Sebald, A., & Stepney, S. (2015). Heterotic computing: Past, present and future. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *373*, 20140225.

Kersten, L. (2017). A mechanistic account of wide computationalism. *Review of Philosophy and Psychology*, *8*, 501–517.

Law, E., & von Ahn, L. (2011). Human computation. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *5*, 1–121.

Lindblom, J., & Ziemke, T. (2003). Social situatedness of natural and artificial intelligence: Vygotsky and beyond. *Adaptive Behavior*, *11*, 79–96.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., & van den Berg, J. (2008). Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *Monthly Notices of the Royal Astronomical Society*, *389*, 1179–1189.

Menary, R. (2010). Introduction: The extended mind in focus. In R. Menary (Ed.), *The Extended Mind*. Cambridge, Massachusetts, USA: MIT Press.

Michelucci, P. (Ed.) (2013). *Handbook of Human Computation*. New York, New York, USA: Springer.

Michelucci, P., & Dickinson, J. L. (2016). The power of crowds. *Science*, *351*, 32–33.

Miller, R. C., Little, G., Bernstein, M., Bigham, J. P., Chilton, L. B., Goldman, M., Horton, J. J., & Nayak, R. (2010). Heads in the cloud. *XRDS: Crossroads*, *17*, 27–31.

Minder, P., & Bernstein, A. (2012). CrowdLang: A programming language for the systematic exploration of human computation systems. In *4th International Conference on Social Informatics*. Lausanne, Switzerland.

Modha, D. S., Ananthanarayanan, R., Esser, S. K., Ndirango, A., Sherbondy, A. J., & Singh, R. (2011). Cognitive computing. *Communications of the ACM*, *54*, 62–71.

Mohedano, E., Healy, G., McGuinness, K., Giró-i Nieto, X., O'Connor, N. E., & Smeaton, A. F. (2014). Object segmentation in images using EEG signals. In *22nd ACM International Conference on Multimedia*. Orlando, Florida, USA.

Murray, D. G., Yoneki, E., Crowcroft, J., & Hand, S. (2010). The case for crowd computing. In *ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds*. New Delhi, India.

Pizzi, R. M. R., Rossetti, D., Cino, G., Marino, D., Vescovi, A. L., & Baer, W. (2009). A cultured human neural network operates a robotic actuator. *Biosystems*, *95*, 137–144.

Polger, T. W. (2009). Computational functionalism. In J. Symons, & P. Calvo (Eds.), *The Routledge Companion to Philosophy of Psychology*. Abingdon, England, UK: Routledge.

Roberts, T. (2012). You do the maths: Rules, extension, and cognitive responsibility. *Philosophical Explorations*, *15*, 133–145.

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

Rowlands, M. (2009). Extended cognition and the mark of the cognitive. *Philosophical Psychology*, *22*, 1–19.

Ruaro, M. E., Bonifazi, P., & Torre, V. (2005). Toward the neurocomputer: Image processing and pattern recognition with neuronal cultures. *IEEE Transactions on Biomedical Engineering*, *52*, 371–383.

Schall, D. (2013). Service oriented protocols for human computation. In P. Michelucci (Ed.), *Handbook of Human Computation*. New York, New York, USA: Springer.

Schall, D., Truong, H.-L., & Dustdar, S. (2008). Unifying human and software services in Web-scale collaborations. *IEEE Internet Computing*, *12*, 62–68.

Simon, H. A. (1996). *The Sciences of the Artificial*. (3rd ed.). Cambridge, Massachusetts, USA: MIT Press.

Smart, P. R. (in press). Emerging digital technologies: Implications for extended conceptions of cognition and knowledge. In A. J. Carter, A. Clark, J. Kallestrup, O. S. Palermos, & D. Pritchard (Eds.), *Extended Epistemology*. Oxford, UK: Oxford University Press.

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*. Hershey, Pennsylvania, USA: IGI Global.

Smart, P. R., Heersmink, R., & Clowes, R. W. (2017). The cognitive ecology of the Internet. In S. J. Cowley, & F. Vallée-Tourangeau (Eds.), *Cognition Beyond the Brain: Computation, Interactivity and Human Artifice*. Cham, Switzerland: Springer International Publishing. (2nd ed.).

Staley, D. J. (2014). *Brain, Mind and Internet: A Deep History and Future*. Basingstoke, England, UK: Palgrave Macmillan.

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

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

Wang, Y., Jiang, L., Cai, B., Wang, Y., Zhang, S., & Zheng, X. (2015). A closed-loop system for rapid face retrieval by combining EEG and computer vision. In *7th International IEEE/EMBS Conference on Neural Engineering*. Montpellier, France.

Weld, D. S., Lin, C. H., & Bragg, J. (2015). Artificial intelligence and collective intelligence. In T. W. Malone, & M. S. Bernstein (Eds.), *Handbook of Collective Intelligence*. Cambridge, Massachusetts, USA: MIT Press.

Wheeler, M., & Clark, A. (2008). Culture, embodiment and genes: Unravelling the triple helix. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*, 3563–3575.

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

Wilson, R. A., & Clark, A. (2009). Situated cognition: Letting nature take its course. In P. Robbins, & M. Aydede (Eds.), *The Cambridge Handbook of Situated Cognition*. Cambridge, UK: Cambridge University Press.

Woods, D. D. (1985). Cognitive technologies: The design of joint human–machine cognitive systems. *AI Magazine*, *6*, 86–92.