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 the idea that human agents can be seen as part of the physical fabric that realizes the cognitive capabilities of a machine-based system. 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 functional role of the human social environment in realizing novel forms of intelligent system. The present paper provides an initial defence of the thesis of human-extended machine cognition. In addition, the paper attempts to show how the notion 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 our attempt to understand the mechanisms underlying human intelligence, the human brain has been a rather obvious focus of interest and attention for the

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

sciences of the mind. In fact, given the way in which 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 (at least from the perspective of mechanistic realization) is to be found solely in the neural realm—that the point source of intelligent thought and action is located inside the head of individual 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.

An alternative vision of the mechanistic underpinnings of human intelligence is provided by a philosophical position known as active externalism [1, 2]. In contrast to the neurocentric view of human intelligence, advocates of active externalism propose that the machinery of the human mind is not to be found *solely* within the inner, 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 and include a range of non-neural (and even non-biological) elements. Active externalism thus provides us with an *extended* view of human cognition: a vision 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 crucial interest and importance to cognitive science. In our quest to understand the material bases of intelligent thought and action, it is clearly 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 [3]. If our attention is focused on only one part of what is, in effect, a larger mechanistically-relevant matrix, then we face the risk that an important array of explanatorily-relevant forces and factors may end up falling beyond our field of view. This is almost certainly likely to undermine our ability to develop a mechanistically-grounded

scientific understanding of human cognitive phenomena.

Active externalism, it should be clear, is a theoretical approach that comes with a host of methodological implications. For inasmuch as we accept the basic tenets of the active externalist position, then it seems our attempt to understand human intelligence will need to focus on more than just the biological brain; in addition, to the biological brain, we will need to broaden our scientific remit to include those elements of the extra-neural and extra-organismic environment that are of causal relevance to human cognitive states and processes. This evinces the need for an extended—or at least a situated [4]—approach to human cognition. Such an approach is all the more important if it is the extended character of human cognition that helps to some of its most distinctive features (e.g., the features that make it unique relative to other forms of terrestrial intelligence) [see 5].

But it is not just our approach to human intelligence that is affected by claims regarding the extended character of human cognition; active externalism is also 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, 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 also a sense, however, in which active externalism may be relevant to AI, irrespective of whether or not claims about the extended character of human cognition should turn out to be true. It may be the case, for example, that an extended perspective provides insight into how an array of difficult or intractable computational problems can be solved by engaging in cognitively-potent forms of causal commerce with an array of materially-heterogeneous resources. Such solutions may yield a range of benefits in terms of (e.g.,) the temporal and energetic costs of particular forms of information processing [see 6]. The result is what might be called an extended—or at least a situated—approach to AI. The aim, in this case, is to develop a better understanding of the mechanisms that allow an AI system to press a range of environmental resources into useful

cognitive service.

In the present paper, I attempt to extend the conventional focus of active externalist theorizing by developing and defending an extended approach to AI[1]. In particular, I claim that the notion of extended realization bases for intelligent thought and action is just as applicable to the attempt to synthesize intelligent systems, as it is to the analysis of their natural counterparts. Beyond this somewhat philosophically-oriented objective, however, I also want to suggest that when we look at the environment in which specific instances of machine intelligence are situated, we often encounter a state-of-affairs in which one or more human agents emerge as constituent elements of the physical machinery that realizes machine-based cognizing. This is what I call 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 states and processes.

The notion of HEMC, 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 an opportunity for bidirectional forms of cognitive extension. If we accept the idea that a machine-

[1] The application of active externalist theorizing to AI systems has been the focus of previous work. Jonker [6], for example, provides a useful overview of the 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 extended cognition that involves the cognitive incorporation of human agents. This is an idea that has not, to my knowledge, been presented in the philosophy of mind, cognitive science and AI literatures.

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

The main aim of the current paper is to provide an initial defence of the HEMC thesis. I also attempt to show that HEMC is more than just a topic of theoretical interest and speculation. In particular, I seek 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 enquiry.

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 can (ever?)[2] be glossed 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

[2]For present purposes, I will assume, in accord with functionalist intuitions about the mind [7], 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.

> of some form of machine-based information processing (extended or otherwise) as a genuine case of machine cognition?"

This, it should be clear, is an important question. If we cannot identify a set of conditions that enable us to reliably distinguish cognitive from non-cognitive phenomena, then the HEMC concept seems to be of nugatory scientific and philosophical value. Ideally, what is required, here, is an understanding of what Adams and Aizawa [8] dub the 'mark of the cognitive': the set of features that enable us to distinguish cognitive phenomena from their non-cognitive counterparts. Only in the light of this understanding will we be able to determine whether some form of machine-based processing (extended or otherwise) should be regarded as a genuine instance of machine-based cognizing.

As is perhaps clear from the very fact that this 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" [8, 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 machine-based systems. There are, however, a number of reasons to doubt the integrity of the claims made by Adams and Aizawa. Clark [9, 10, 2], for example, has mounted a robust philosophical defence of the notion of extended cognition in the wake of Adams and Aizawa's critique. Such defences, I suggest, are just as applicable to machine cognition as they are to extended forms of human cognition.

Another view of cognition is proposed by Rowlands [11]. He suggests that a cognitive process should be seen as a process that is "(i) required for the ac-

complishing 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, it is unclear why we cannot resort to a similar ostensive strategy when it comes to the individuation of cognitive processes.

Finally, Clark [10] has suggested that the notion of a cognitive process might be "best unpacked as the notion of a process that supports certain kinds of behaviour (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 behaviour. He suggests that "To identify cognitive processes as those processes, however many and varied, that support intelligent behavior may be the best we can do" [10, p. 93].

I will not attempt to adjudicate between these different approaches to cognition. What matters, for present purposes, is that none of these different ways of conceptualizing cognition are incompatible with the basic possibility of machine cognition[3]. Clark and Rowlands are both happy to place considerable emphasis

[3]One exception, perhaps, is the issue of non-derived contents (or intrinsic intentionality), as discussed by Adams and Aizawa [8]. 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 the case of HEMC seems to fit the bill, even if the case of

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 behaviour). Rowlands, additionally, makes explicit reference to the notion of information processing, namely, that a cognitive process is one that involves the processing of information. 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 behaviour. Neither of these conditions, it should be clear, pose any problem for the claim that some kinds of machine-based system can possess cognitive properties. Indeed, the current interest in 'cognitive computing systems' [12, 13, 14] is testament to the fact that at least some forms of intelligent system can be viewed in cognitive terms.

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 [1], as part of their seminal work on extended cognition. 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" [1, p. 8].

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

non-extended machine cognition does 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.

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**
> 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 occured within the head of an individual human agent.

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 (neuro-)cognitive processing. If the answer to this question is 'yes' (in the counterfactual case), then why should we revise our judgements about the cognitive status of the relevant process just because it is realized (in the actual case) by some alternative nexus of material elements? As should be clear, the Parity Principle for Machine Cognition is intended to shield our evaluative efforts from any form of bio-centric prejudice or bias.

3. Extended Cognition

The term 'extended cognition' identifies an important and influential body of work within the philosophy of mind 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-biological, extra-organismic resources, such as

physical artefacts and technological devices [2, 1]. Extended cognition is thus a claim about the nature of the realization base for some episode of intelligent behaviour. 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 an array of materially-heterogeneous[4] resources, which includes both the usual biological suspects (i.e., the brain and body) and a set of non-biological elements (e.g., elements 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 would thus seek to explain how we first derive some symbolic encoding of the visual (or auditory) input corresponding to the two numbers, and 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 explanatory 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 is profoundly simplified. In place of purely inner, environmentally-decoupled, computational operations we now see a pattern

[4]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 [15]. There is, in this case, no heterogeneity in the underlying realizing mechanisms, since all the mechanisms are located in the biological realm.

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 [16], is a nice example of what might be called environmentally-extended computation or 'wide computationalism' [see 17]. 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 latter, world-involving account is that the extra-neural resources are 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 that of the central executive. What the notion of extended cognition gives us, therefore, is the idea of extra-neural and extra-corporeal resources serving as part of the explanatorily-relevant physical fabric that accounts for our success 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. And, in cases where the target phenomenon is 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, in the case of extended cognition, is that we confront a set of distinct 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 highlights the fact that we should think of an extended cognitive organization as a form of 'multi-modal'[5] network system [19]. We thus encounter a case of extended cognizing when the information processing circuits that realize a particular cognitive process (or task) 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 [20, 21]. It is thus the 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 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 notion of HEMC.

4. Human-Extended Machine Cognition

Now that we have a better understanding of what is meant by the notions of 'machine cognition' and 'extended cognition', it is time to turn our attention to the concept of HEMC. Perhaps the best way of creeping up on the notion of HEMC 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 (or system). 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

[5] 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 18].

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 individual, we are instead encouraged to see the human agent as a resource that is incorporated into the cognitive processing routines of a synthetic system. A useful way of thinking about HEMC is thus to see it as a form of active externalism (or 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 towards the technological elements of the system: it is the 'cognitive' capabilities of the technological elements that we see as being extended (and perhaps enhanced) as a result of the specific form of techno-biological bonding that occurs.

The HEMC concept, it should be clear, is a concept that is firmly rooted in contemporary philosophical debates regarding 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 agents. In other words, HEMC encourages us to take an extended view of machine cognition. This idea is important for a number of reasons. Firstly, the HEMC concept is able to highlight the philosophical relevance of recent work concerning human computation systems [22, 23] and crowdsourcing techniques [24]. A common focus of interest in these areas is the extent to which human agents can be actively incorporated into computational tasks that are managed, monitored and maintained by machine-based systems. A virtue of the HEMC concept is thus the way that it enables us to bring together a num-

ber of disparate bodies of work under a common conceptual umbrella. Work on human computation, which, for the most part, has been undertaken as part of the discipline of computer science, has typically proceeded independently of work in the 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 bridging work in these areas. The resulting intellectual payoff is, I suggest, significant. In particular, by taking the notion of HEMC seriously, a number of important issues are brought to light. These include issues of complementarity and hybridization, which are a prominent focus of attention for the proponents of extended cognition [25, 26]. With the notion of HEMC to hand, we now have a useful means of juxtaposing these philosophical debates with work on complementarity and hybridization in computer science. This includes work on complementary computing [27], mixed-initiative systems [28], interactive machine learning [29], symbiotic interaction [30], and heterotic computing [31].

A second reason why the notion of HEMC is important is because of its potential to broaden the philosophical debate relating to extended cognition. A key example of this occurs in respect of what is called the Hypothesis of Organism-Centred Cognition (HOC) [2]. 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" [2, p. 139].

The HOC thus situates a biological organism at the heart of an extended cognitive system. This, in fact, 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 32]. Hutchins [33], 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 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 looks to be overly bio-centric. A revised version of the HOC should, I suggest, 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, we are provided with a potentially useful way of thinking about 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 5][6], then perhaps the route to synthetic forms of human-level intelligence lies not so much in the development of new kinds of computational system, as the development of systems that are able to press maximal cognitive benefit from the environments in which they are materially embedded.

[6]Clark [5], 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).

5. When is Machine Cognition Extended?

Assuming that we accept the basic notion of HEMC, an important question to ask is when a particular instance of machine-based processing should be regarded as a *bona fide* case of HEMC? In other words, what are the conditions under which an episode of machine-based processing should (or should not) be seen to be materially-extended? The answer to this question, it should be clear, is one that relies on our ability to demarcate the typical boundaries of a cognitive agent, such as a human individual, a robot or a computational system. It also depends on our ability to discern when the physical machinery that realizes a cognitive process extends beyond the physical borders of the target agent. The problem is that not every form of human-machine interaction counts as a genuine case of HEMC, so we need to specify the conditions under which HEMC occurs. When, exactly, should some form of human-machine (or, perhaps more appropriately, machine-human) interaction be cast as a genuine case of HEMC?

In answering this question, I suggest that we should pay attention to the following criteria. We encounter a case of HEMC, I suggest, when:

1. we are confronted with a process (or task) that we are prepared to accept as a cognitive process (or task),
2. the process (or task) in question is realized by a mixture of human and machine elements (i.e., the realization base for the process is bio-technologically hybrid), and
3. the machine 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 machine elements that are 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 8, 10]. Given that we have already discussed this issue at some

length (see Section 2), let us turn our attention towards 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 be smeared across a smorgasbord of materially heterogeneous resources. 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 processing. 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 advanced form of AI. Presumably, in this case you would 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 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 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 cognitive computing system. Functional parity is thus (at least) 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 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 also part of realization base for what are, in effect, the same performances?

If this example doesn't grab you, perhaps it will help to consider a more concrete case involving an existing system that relies on the contributions of a large number of human volunteers to perform a cognitive task. Let us therefore turn our attention to a system called 'Galaxy Zoo' [34]. Galaxy Zoo is a 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. [34], 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 35], the services of the human agents are no longer required. In this case, *all* the relevant classification processes are translocated to the technological realm and are thus subject to technological realization. Now, inasmuch as we accept that the galaxy classification task is a cognitive task, then there seems little reason to reject the idea that the automated (computerized) version of the task should not be seen in the same way as the task performed by the human-machine ensemble—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, and that the process in question was realized by a mixture of human

and machine elements.

At this point, it may help to consider another version of the Parity Principle that was discussed in Section 2. The basic idea of the Parity Principle, recall, is to help us ignore the traditional biological borders of skin and skull in determining whether some part of the physical world should count as part of the supervenience base for some cognitive phenomenon of interest. A variant of the Parity Principle—the Reverse Parity Principle—can, I suggest, be used to guide our intuitions as to when we confront a genuine case of HEMC:

> **Reverse Parity Principle**
>
> If, as we confront some task, a human agent functions as a process, which were it undertaken solely by a machine-based processor, we would have no hesitation in recognizing it as part of the cognitive processing undertaken by the machine, then the human agent is (for that time) part of the cognitive process.

What the Reverse Parity Principle is asking us to do, here, is to imagine a situation in which the functional contributions of the human agent are, contrary to all the facts, being implemented by a technological component. If, in the counterfactual case, we are content to see the non-biological elements as part of the information processing economy of some larger computational organization, then why should we demur from the conclusion that the human agent forms part of the realization base for intelligent performances in the actual, real-world case?

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 concept. 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 endeavours regarding the possession of specific

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 [36, 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 artefacts [see 37]. 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 controlling the flow of information, 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 notions of cognitive agency when it comes to our understanding of HEMC systems. 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 [2] 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 extended machinery of its own thought and reason" (p. 123).

What the vision of HEMC requires is that we reverse the usual order of things and see the machine elements as playing a role akin to that of the human agent in the case of conventional (human-centered) forms of extended cognition. 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 its cognitively-relevant performances. As is noted by Clark [38, 39], a useful way of thinking about these processes of creation, coordination and control is to see them as meta-cognitive processes; i.e., as processes that embody knowledge concerning, among other things, the relationship between different structural configurations and task-relevant information processing outcomes. The function of these meta-cognitive processes (in both the human and machine cases) is, I suggest, 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 system. We will encounter some specific examples of these forms of machine-based 'meta-cognitive control' in later sections (see Section 8 and Section 9).

6. Neural Wideware

The term 'wideware' is a term used by proponents of extended cognition to refer to the physical elements that comprise an extended cognitive circuit [40]. 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 is, of course, the elements that lie outwith the neural realm. In discussing the notion of wideware, for example, Clark [40] 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 obliges us to revise our traditional understanding of the notion of wideware and give serious thought to the possibility of 'neural wideware', a specific form of wideware in which it is the biological brain of a human individual (or collections thereof) that are playing an explanatorily-significant role in accounting for the cognitive capabilities of a machine-based system.

In order to help us develop a better understanding what is being proposed here, consider Clickworkers [41], one of the early demonstrations of the power and potential of crowdsourcing techniques for what is now known as citizen science [see 42]. 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 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 (i.e., the outline of craters). The images would then be subject to automated image segmentation routines that decompose the original image array into a set of image tiles[7]. 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. [44]. 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 the machine attempts to reconcile the recordings with the results obtained from other human subjects, as well the outputs of its own image processing routines [see 43][8].

[7]Both of these processes are, in fact, demonstrated by recent attempts to apply image processing techniques to the crater mapping task [see 43]

[8]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

The resulting (hypothetical) system is, I suggest, a genuine case of HEMC, one in which we see 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 neurocomputational 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. In order to help us see this, we can invoke the conceptual services of the Reverse Parity Principle (see Section 5). Suppose, for example, 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 realizes the crater mapping task. This seems unlikely. But once we accept this point, there seems little reason to reject the claim that the brains of the human agents in the bio-technologically hybrid should be seen as part of the material fabric that realizes the target process.

HEMC thus opens the door to claims about neural wideware—the idea that neurological elements can, on occasion, emerge as constituent elements of an extended cognitive circuit[9]. But notice that this idea is one that conflicts with the original, Clarkian notion of wideware presented above. Clark [40] suggests that the notion of wideware excludes the possibility of brain-based forms of realization; however, the notion of neural wideware, which emerges as a specialization of the wideware concept, *does* allow for this particular form of realization. Here, then, we can begin to see the philosophical relevance and importance of

---

of enhancing the current state-of-the-art in image processing [45, 46].

[9]Such claims, it seems to me, make perfect sense in the wake of research concerning so-called neurocomputers. Here, the computational capacities of some neural resource (typically an *in vitro* neural preparation) are used to assist with the information processing tasks assigned to some larger system (e.g., a robotic agent) [47, 48, 49, 50].

the HEMC concept. In this case, the HEMC concept helps to reveal a potential problem with one of the claims associated with the active externalist position.

7. Human Computation

The HEMC concept is, I suggest, a concept that is widely applicable to a substantial body of work that forms part of the theoretical, scientific and engineering agenda of contemporary computer science. Crucially, the advent of the Internet, coupled with recent developments in mobile and portable computing technology, has profoundly altered the kinds of opportunities that human agents have to interact with machine-based systems. This has led to a preoccupation with the kinds of ways in which human agents can be factored into episodes of machine-based processing. Evidence of such a preoccupation can be found in the literature on active learning systems [51], citizen science systems [42], online computer games [52], interactive evolutionary computation systems [53], hybrid human-machine systems [54], social machines [55, 56], cognitive computing systems [14] and human computation systems [22, 23]. In one way or another, all these various areas of research express an interest in tapping into the cognitive capabilities of human agents, typically for the purposes of tackling problems that are beyond the computational reach of contemporary AI systems.

The HEMC concept is of particular relevance to work that goes under the heading of 'human computation'. To see this, we need only consider some of the ways in which the notion of human computation has been defined within the computer science literature [see 57]. Law and von Ahn [22], for example, suggest that:

> "Human computation is a new and evolving research area that centers around harnessing human intelligence to solve computational problems...that are beyond the scope of existing Artificial Intelligence (AI) algorithms" (p. xiii).

Here, we encounter an explicit commitment to the idea that human agents can be relied on to supplement (or even supplant) machine-based processing.

Other definitions go even further in terms of establishing a link with the notion of HEMC. Some definitions, for example, intimate at the possibility of human agents being incorporated, merged or integrated into computational routines [e.g., 58]. On this basis, we might be inclined to think that the extensional projection of the terms 'human computation system' and 'HEMC system' are in perfect alignment. There are, however, a number of points of difference between the notions of human computation and HEMC. An appreciation of these differences is important, if only because it helps to guard against claims of conceptual redundancy (i.e., it prevents HEMC being seen as equivalent to human computation and therefore redundant from a conceptual perspective)[10].

The first thing to note, here, is that although the term 'human computation' is typically applied to systems featuring a combination of human and machine processing, there is no reason why the term cannot be applied to systems that consist solely of human agents. Indeed, as is noted by Grier [59] the historical precursors of today's more technologically-advanced forms of human computation date back to at least the 18th century. In these early cases, the relevant systems lacked any form of mechanized information processing. This marks an important difference between the notions of HEMC and human computation, for at the heart of the HEMC concept is the idea that the machine-based components of a system should be *actively* engaged in the realization of cognitive phenomena.

A second point of interest is inspired by the aforementioned notion of neural wideware (see Section 6). Neural wideware, recall, is based on the idea that the neural resources of a human agent could be exploited for the purposes of completing a cognitive task. In the case of the rapid serial visual presentation paradigm described by Huang et al. [44], for example, EEG recordings were used

[10] An obvious difference between the two terms is the fact that human computation is concerned with computational processes that or may not qualify as cognitive processes. Conversely, HEMC is concerned with cognitive processes that may or may not qualify as computational processes.

to draw on the computational capabilities of the biological brain as a means of servicing the information processing demands of an image processing task. We can, however, move beyond the specific case of neural wideware and broaden the scope of the concept to include other kinds of physiological processes. We can thus consider techniques that rely on the use of eye movements [60], galvanic skin responses [61], or facial expressions [62] as a means of incorporating human agents (or their constituent physiological processes) into machine-based processing routines. Crucially, however, the kind of incorporation that occurs in these sorts of cases seems to be radically different to that envisioned by the proponents of human computation. When it comes to human computation, the focus of attention is typically a collection of overt physical actions that are both deliberate and oriented to the task in question. Human computation theorists thus tend to see the entire human agent as something like a computational processor that interacts with a set of bio-externally situated symbolic representations and manipulates them according to some predefined set of instructions. According to this vision, it is at best unclear as to whether we should count the case of neural (or perhaps physiological) wideware as compatible with the notion of human computation. In fact, I suggest that this is unlikely to be the case. The reason for this is that human computation theorists tend to view the contributions of the human agent from a computational perspective. In other words, they see the actions of the human agent as constituting a form of computational processing. However, when it comes to cases of physiological wideware (consider, for example, the case of galvanic skin responses), it seems far from clear that the biological components are implementing a process that deserves to be glossed in computational terms. Much of course depends on how we choose to define the notion of 'computation' [see 63]. The main point, for present purposes, however, is that there is nothing in the HEMC concept that requires the contributions of the human agent to be afforded a computational interpretation. The notion of HEMC is thus able to accommodate cases where overt human actions are seen to play a substantive role in the mechanistic realization of an extended cognitive routine (as is often the case in human computation). However, it is also able to

accommodate cases where the human agent is seen to play something of a more 'passive' role (as in the cases involving physiological wideware).

A final point of difference between the notions of HEMC and human computation concerns the role of conscious awareness. In discussing the role of human agents in collecting bodies of environmental data, for example, Law and von Ahn [22] suggest that:

> "...a participatory sensing project where humans actively decide where and when to collect information can be considered human computation; on the other hand, it is not human computation if the participants are merely the sensor carriers, with no *conscious* role in determining the outcome of the computation" (p. 4) [original emphasis].

Conscious awareness thus seems to play an important role in at least some cases of human computation. The same, however, is not the case for HEMC. In the case of HEMC, there is no requirement for the human individual to be aware of the kind of role they are playing in a cognitive system. Indeed, there is no requirement for the human agent to even be aware that they are a part of that system; i.e., that they are a cog in a much larger cognitive machine. To my mind, the fact that there is no awareness of the larger task, or conscious intent to supply a technological system with some requisite body of task-relevant information, in no way influences claims regarding the extended or non-extended nature of the larger system. Does it matter that, in the long multiplication case discussed in Section 3, the pen and pencil resources have no conscious role in determining the mathematical outcome of the environmentally-extended process? Or, if you find the multiplication case problematic, what about the case of an individual neuron that is involved in realizing some internally-situated form of multiplicative activity? Does it matter that the individual neuron has no awareness of its role in determining the outcome of the purely internal cognitive process? And does this lack of conscious intent, in some way, undermine the status of the neuron as an intrinsic part of the physical fabric that realizes the

relevant cognitive process? The answer to these questions is surely a resounding ‘no’: the ability of an individual component of a cognitive system to engage in some form of deliberate decision making or exhibit some form of conscious intent in no way undermines the status of that component as a candidate part of the physical machinery that realizes a cognitive state or process.

It should, by now, be clear that the notions of human computation and HEMC are not equivalent. This does not mean, however, that the HEMC concept is irrelevant to work on human computation systems. For even if we accept the constraints relating to conscious awareness and overt physical action, this does not mean that the HEMC concept cannot be applied to at least some human computation systems: the HEMC concept, recall, is simply agnostic with respect to the proposed constraints; it does not exclude the possibility that some HEMC systems will be ones where the proposed constraints are, in fact, fully satisfied.

8. Humans in the Loop

One response to the HEMC concept is to see it as identifying a potentially interesting focus for future research efforts in AI, some of which may yield a novel class of intelligent system. This is certainly one way to view the HEMC concept. The value of the HEMC concept is, however, also apparent in the way in which it helps us to view a number of *extant* systems. Consider, for example, a system described by Branson et al. [64, 65], which relies on a combination of machine vision algorithms and human 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 both humans and machines to perform. When presented with the photograph shown in Figure 1, 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 an Indigo Bunting (*Passerina cyanea*). Similarly, the task of species identification is one

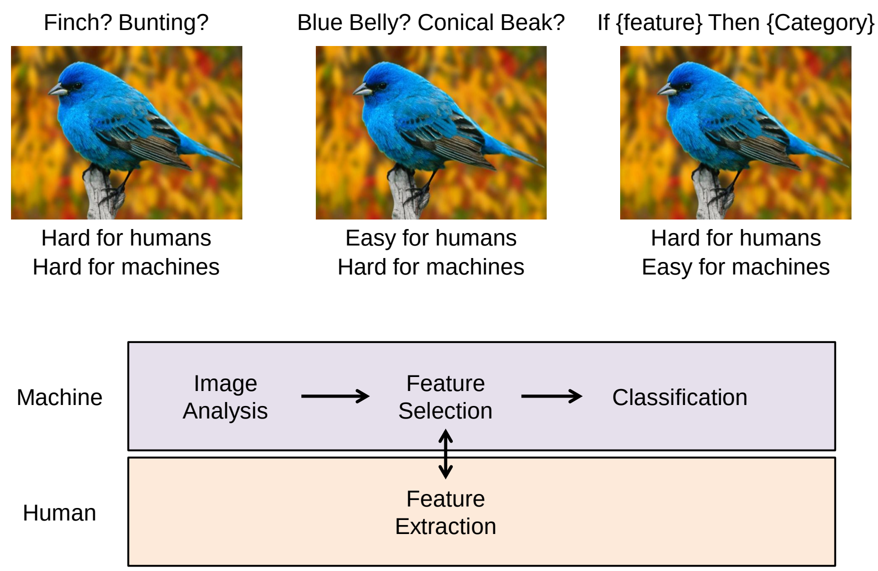

Figure 1: Identifying the species of bird shown in a photographic image is a task that is difficult for both humans and machines. The task can, however, be broken down into a series of smaller steps and assigned to either the human or machine components of a larger system. The result is a hybrid knowledge-based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

that is at least challenging for 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 (see Figure 1). For example, one of the steps in the species identification task relates to the extraction of specific features. These include features relating to (e.g.) 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 sub-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 species-related responses. The selection of appropriate feature-oriented questions is thus something of a knowledge-intensive task in its own right, one which requires an ability to calculate the relative optimality (in an information theoretic sense) of different 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 system (see Figure 1).

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 genuine case 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 judgements 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) being used to guide our intuitions as to the cognitive status of the focal task.

Secondly, there can be little doubt that the realization base for the cognitive performances are, in the majority of cases[11], bio-technologically extended. If

[11] In some cases, the task was solved *solely* by the machine-based components of the system.

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 an intrinsic part of the physical fabric that realizes the target task. In this case, we find ourselves invoking the Reverse Parity Principle, as discussed in Section 5.

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 will get presented to the human community, and it is the technological elements that determine *when* these particular questions will be 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'. In this case, the nature of the informational contact that is established with the human user community is determined dynamically (at runtime) as a specific result of the *machine's* own appreciation of the current problem state. The machine thus selects the questions that are likely to be the most informative, based on the results of previous information processing (which may be of either the extended or non-extended variety). The machine also plays a role in 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 a more difficult case, 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 may be

---

This occurred in cases where the machine vision algorithms were able to process the target image and yield an acceptable result.

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. Even in this situation, however, 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 agents.

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 66]. 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 system.

The result of all this is that the HEMC-related criteria presented in Section 5 are fully 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. [64] as a genuine real-world example of HEMC.

9. Ecological Assembly

One of the interesting features of Branson et al.'s system is that it exhibits of form of contingent connectivity. Connections with the human community are

thus 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: 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 relating to 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) [2]. 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" [2, p. 13]. The image that emerges, here, is one in which the human biological agent can be seen to play an active role in the creation of the world-involving informational circuits that subsequently realizes some aspect of her cognitive processing.

It is here, I suggest, that we can 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, in this case, 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 an environment that consists mostly of us!

Interestingly, once we view the PEA from the perspective of HEMC, the

philosophical and cognitive scientific significance of a number of other strands of computer science research are brought into sharper focus. These include efforts 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 the 'human cloud' [67]. 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 [68]. 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, 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 to 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-technological entanglement and integration [69, 70], the extension of traditional Web service description languages to accommodate the possibility of human involvement [69], and the emergence of programming frameworks that are specifically intended to deliver "complex computation systems incorporating large crowds of networked humans and machines" [71, p. 124]. Progress in these various areas of research are 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 organizations. 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 HEMC-based systems.

10. 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 capabilities of machine-based systems. This proposal is both novel and undoubtedly contentious. It is, however, an idea that is both appropriate and relevant, especially when one considers the current interest in hybrid systems [54], human computation [23], heterotic computing [31], crowd computing [72], and collective intelligence [73, 74].

In the present paper, I have attempted to provide an initial defence 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 sorts of 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 8). As a result of this analysis, I demonstrated that the HEMC concept is not just a focus of theoretical interest for the philosophy community, it is also concept that can be applied to a number of existing systems that are the current focus of attention for the computer science community.

The HEMC concept has a number of important implications that are spread across an array of disciplines. Firstly, the HEMC thesis is of significant 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 concerning

extended cognition. We saw, for example, that the HEMC thesis poses a challenge to the HOC (see Section 4), as well as the Clarkian notion of wideware (see Section 6). Given that these concepts are widely accepted within the philosophical community, it will be important to ascertain the extent to which these concepts should be revised in the wake of the HEMC thesis. It will also be important to determine the impact of these revisionary efforts on our understanding of extended cognition.

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 (see Section 7) and the 'human cloud' (see Section 9). We are also able to 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 [75, 76, 77]. When we look at the Internet through the conceptual lens of HEMC, however, 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 see the roots of human intelligence as tied to an ability to exploit the features of the extra-organismic environment [5], 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, 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 the global human population.

Acknowledgements

This work is supported under SOCIAM: The Theory and Practice of Social Machines. The SOCIAM Project is funded by the UK Engineering and Physical Sciences Research Council (EPSRC) under grant number EP/J017728/1 and comprises the Universities of Southampton, Oxford and Edinburgh.

References

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

[2] A. Clark, Supersizing the Mind: Embodiment, Action, and Cognitive Extension, Oxford University Press, New York, New York, USA, 2008.

[3] C. Craver, J. Tabery, Mechanisms in science, in: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, Spring 2016 Edition, Stanford University, Stanford, California, USA, 2016.

[4] P. Robbins, M. Aydede (Eds.), The Cambridge Handbook of Situated Cognition, Cambridge University Press, New York, New York, USA, 2009.

[5] A. Clark, Natural-Born Cyborgs: Minds, Technologies and the Future of Human Intelligence, Oxford University Press, Oxford, UK, 2003.

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

[7] T. W. Polger, Computational functionalism, in: J. Symons, P. Calvo (Eds.), The Routledge Companion to Philosophy of Psychology, Routledge, Abingdon, England, UK, 2009.

[8] F. Adams, K. Aizawa, Defending the bounds of cognition, in: R. Menary (Ed.), The Extended Mind, MIT Press, Cambridge, Massachusetts, USA, 2010.

[9] A. Clark, Memento's revenge: The extended mind, extended, in: R. Menary (Ed.), The Extended Mind, MIT Press, Cambridge, Massachusetts, USA, 2010.

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

[11] M. Rowlands, Body Language: Representation in Action, MIT Press, Cambridge, Massachusetts, USA, 2006.

[12] J. S. Hurwitz, M. Kaufman, A. Bowles, Cognitive Computing and Big Data Analytics, John Wiley & Sons, Inc., Indianapolis, Indiana, USA, 2015.

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

[14] D. S. Modha, R. Ananthanarayanan, S. K. Esser, A. Ndirango, A. J. Sherbondy, R. Singh, Cognitive computing, Communications of the ACM 54 (8) (2011) 62–71.

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

[16] R. A. Wilson, A. Clark, Situated cognition: Letting nature take its course, in: P. Robbins, M. Aydede (Eds.), Cambridge Handbook of Situated Cognition, Cambridge University Press, Cambridge, UK, 2009.

[17] R. A. Wilson, Wide computationalism, Mind 103 (411) (1994) 351–372.

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

[19] P. R. Smart, P. C. Engelbrecht, D. Braines, M. Strub, C. Giammanco, The network-extended mind, in: D. Verma (Ed.), Network Science for Military

Coalition Operations: Information Extraction and Interaction, IGI Global, Hershey, Pennsylvania, USA, 2010.

[20] R. Clowes, Thinking in the cloud: The cognitive incorporation of cloud-based technology, Philosophy & Technology 28 (2) (2015) 261–296.

[21] T. Roberts, You do the maths: Rules, extension, and cognitive responsibility, Philosophical Explorations 15 (2) (2012) 133–145.

[22] E. Law, L. von Ahn, Human computation, Synthesis Lectures on Artificial Intelligence and Machine Learning 5 (3) (2011) 1–121.

[23] P. Michelucci (Ed.), Handbook of Human Computation, Springer, New York, New York, USA, 2013.

[24] P. Michelucci, J. L. Dickinson, The power of crowds, Science 351 (6268) (2016) 32–33.

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

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

[27] A. Kapoor, D. Tan, P. Shenoy, E. Horvitz, 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, 2008.

[28] E. J. Horvitz, Reflections on challenges and promises of mixed-initiative interaction, AI Magazine 28 (2) (2007) 19–22.

[29] J. Fogarty, D. Tan, A. Kapoor, S. Winder, CueFlik: Interactive concept learning in image search, in: SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, 2008.

[30] G. Jacucci, L. Gamberini, J. Freeman, A. Spagnolli (Eds.), Symbiotic Interaction, Springer, London, UK, 2014.

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

[32] A. Clark, Finding the mind, Philosophical Studies 152 (3) (2011) 447–461.

[33] E. Hutchins, Enculturating the supersized mind, Philosophical Studies 152 (3) (2011) 437–446.

[34] C. J. Lintott, K. Schawinski, A. Slosar, K. Land, S. Bamford, D. Thomas, M. J. Raddick, R. C. Nichol, A. Szalay, D. Andreescu, P. Murray, J. van den Berg, Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey, Monthly Notices of the Royal Astronomical Society 389 (3) (2008) 1179–1189.

[35] S. Dieleman, K. W. Willett, J. Dambre, Rotation-invariant convolutional neural networks for galaxy morphology prediction, Monthly Notices of the Royal Astronomical Society 450 (2) (2015) 1441–1459.

[36] A. Clark, Being There: Putting Brain, Body and World Together Again, MIT Press, Cambridge, Massachusetts, USA, 1997.

[37] P. R. Smart, 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 University Press, Oxford, UK, in press.

[38] A. Clark, Surfing Uncertainty: Prediction, Action and the Embodied Mind, Oxford University Press, New York, New York, USA, 2016.

[39] A. Clark, What 'extended me' knows, Synthese 192 (11) (2015) 3757–3775.

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

[41] B. Kanefsky, N. G. Barlow, V. C. Gulick, Can distributed volunteers accomplish massive data analysis tasks, in: 32nd Lunar and Planetary Science Conference, Houston, Texas, USA, 2001.

[42] C. J. Lintott, J. Reed, Human computation in citizen science, in: P. Michelucci (Ed.), Handbook of Human Computation, Springer, New York, New York, USA, 2013.

[43] T. Barata, E. I. Alves, J. Saraiva, P. Pina, Automatic recognition of impact craters on the surface of Mars, in: International Conference of Image Analysis and Recognition, Porto, Portugal, 2004.

[44] Y. Huang, D. Erdogmus, M. Pavel, S. Mathan, K. E. Hild, A framework for rapid visual image search using single-trial brain evoked responses, Neurocomputing 74 (12) (2011) 2041–2051.

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

[46] Y. Wang, L. Jiang, B. Cai, Y. Wang, S. Zhang, X. Zheng, 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, 2015.

[47] M. E. Ruaro, P. Bonifazi, V. Torre, Toward the neurocomputer: Image processing and pattern recognition with neuronal cultures, IEEE Transactions on Biomedical Engineering 52 (3) (2005) 371–383.

[48] T. B. DeMarse, K. P. Dockendorf, Adaptive flight control with living neuronal networks on microelectrode arrays, in: IEEE International Joint Conference on Neural Networks, Montreal, Quebec, Canada, 2005.

[49] T. B. DeMarse, D. A. Wagenaar, A. W. Blau, S. M. Potter, The neurally controlled animat: Biological brains acting with simulated bodies, Autonomous Robots 11 (3) (2001) 305–310.

[50] R. M. R. Pizzi, D. Rossetti, G. Cino, D. Marino, A. L. Vescovi, W. Baer, A cultured human neural network operates a robotic actuator, Biosystems 95 (2) (2009) 137–144.

[51] L. Barrington, D. Turnbull, G. Lanckriet, Game-powered machine learning, Proceedings of the National Academy of Sciences 109 (17) (2012) 6411–6416.

[52] L. von Ahn, Games with a purpose, Computer 39 (6) (2006) 96–98.

[53] J. Secretan, N. Beato, D. B. D Ambrosio, A. Rodriguez, A. Campbell, K. O. Stanley, Picbreeder: Evolving pictures collaboratively online, in: SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, 2008.

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

[55] J. Hendler, T. Berners-Lee, From the Semantic Web to social machines: A research challenge for AI on the World Wide Web, Artificial Intelligence 174 (2010) 156–161.

[56] P. R. Smart, N. R. Shadbolt, Social machines, in: M. Khosrow-Pour (Ed.), Encyclopedia of Information Science and Technology, IGI Global, Hershey, Pennsylvania, USA, 2014.

[57] A. Quinn, B. Bederson, Human computation: A survey and taxonomy of a growing field, in: Annual Conference on Human Factors in Computing Systems (CHI'11), Vancouver, British Columbia, Canada, 2011.

[58] P. Michelucci, Human computation and convergence, in: W. Bainbridge, M. Roco (Eds.), Handbook of Science and Technology Convergence, Springer, New York, New York, USA, 2016.

[59] D. A. Grier, Human computation and divided labour: The precursors of modern crowdsourcing, in: P. Michelucci (Ed.), Handbook of Human Computation, Springer, New York, New York, USA, 2013.

[60] A. Jaimes, J. B. Pelz, T. Grabowski, J. S. Babcock, S.-F. Chang, Using human observer eye movements in automatic image classifiers, in: Human Vision and Electronic Imaging VI, San Jose, California, USA, 2001.

[61] O. Barral, G. Jacucci, Applying physiological computing methods to study psychological, affective and motivational relevance, in: G. Jacucci, L. Gamberini, J. Freeman, A. Spagnolli (Eds.), Symbiotic Interaction, Springer, London, UK, 2014.

[62] I. Arapakis, I. Konstas, J. M. Jose, Using facial expressions and peripheral physiological signals as implicit indicators of topical relevance, in: 17th ACM International Conference on Multimedia, Beijing, China, 2009.

[63] G. Piccinini, Physical Computation: A Mechanistic Account, Oxford University Press, Oxford, UK, 2015.

[64] S. Branson, G. Horn, C. Wah, P. Perona, S. Belongie, The ignorant led by the blind: A hybrid human-machine vision system for fine-grained categorization, International Journal of Computer Vision 108 (1) (2014) 3–29.

[65] S. Branson, C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, S. Belongie, Visual recognition with humans in the loop, in: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 2010.

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

[67] E. Kaganer, E. Carmel, R. Hirschheim, T. Olsen, Managing the human cloud, MIT Sloan Management Review 54 (2) (2013) 23–32.

[68] B. Hayes, Cloud computing, Communications of the ACM 51 (7) (2008) 9–10.

[69] D. Schall, Service oriented protocols for human compuation, in: P. Michelucci (Ed.), Handbook of Human Computation, Springer, New York, New York, USA, 2013.

[70] D. Schall, H.-L. Truong, S. Dustdar, Unifying human and software services in Web-scale collaborations, IEEE Internet Computing 12 (3) (2008) 62–68.

[71] P. Minder, A. Bernstein, CrowdLang: A programming language for the systematic exploration of human computation systems, in: 4th International Conference on Social Informatics, Lausanne, Switzerland, 2012.

[72] D. G. Murray, E. Yoneki, J. Crowcroft, S. Hand, The case for crowd computing, in: ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds, New Delhi, India, 2010.

[73] J. P. Bigham, M. S. Bernstein, E. Adar, Human-computer interaction and collective intelligence, in: T. W. Malone, M. S. Bernstein (Eds.), Handbook of Collective Intelligence, MIT Press, Cambridge, Massachusetts, USA, 2015.

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

[75] P. R. Smart, R. Heersmink, R. Clowes, The cognitive ecology of the Internet, in: S. J. Cowley, F. Vallée-Tourangeau (Eds.), Cognition Beyond the Brain, 2nd Edition, Springer-Verlag, London, England, UK, in press.

[76] D. J. Staley, Brain, Mind and Internet: A Deep History and Future, Palgrave Macmillan, Basingstoke, England, UK, 2014.

[77] N. Carr, The Shallows: How the Internet Is Changing the Way We Think, Read and Remember, Atlantic Books, London, UK, 2010.