

Does the Web Extend the Mind?

Harry Halpin

L'Institut de recherche et d'innovation and W3C/MIT
4 rue Aubry le Boucher
75004 Paris, France
hhalpin@w3.org

ABSTRACT

Under what conditions does the Web count as a part of your own mind? We iterate the conditions upon which cognitive extension and integration can be upheld, and inspect these in light of the Web. We also argue that this ability to integrate the mind into media such as the Web is inherently social, insofar as it involves interaction. Also, there are many cases where external media like the Web are not actually cognitively integrated, but simply serve as a way to co-ordinate intelligent problem-solving via distributed cognition. Yet distributed cognition should not be underestimated, as it is precisely distributed cognition that can solve problems that an individual human may be unable to solve by themselves, and so can serve as a stepping stone to a wider kind of cognitive integration: collective intelligence. Finally, we define collective intelligence as cognitive integration combined with distributed cognition.

Author Keywords

extended mind, philosophy, cognitive integration, Web, distributed cognition, collective intelligence

ACM Classification Keywords

I.2.0.b Artificial Intelligence: Philosophical foundations

General Terms

Cognitive Science; Theory of Mind; Human Factors.

INTRODUCTION

Under what conditions does the Web count as part of your own mind? Is it possible that your mind can extend into the Internet itself? While such a question can be thought of as a mere provocation, the question is deadly serious, and the future of philosophical questions in general, ranging from semantics to ethics, may very well rest upon the answer.

The intuitive answer to whether or not the Web can count as part of the mind often differs depending on the background of those who are asked. The Extended Mind Hypothesis of

Clark and Chalmers state that under certain conditions parts of what would normally be considered the “external environment” outside of the boundaries of the brain and even skin can count as part of the mind [8]. Yet its reception in philosophy of the mind has been somewhat hostile [1] despite a lack of convincing counter-arguments to the Extended Mind Hypothesis itself [7]. Strangely enough, anecdotal evidence seems to show that programmers find it self-evident that the mind can encompass part of the external environment. The inventor of the Web, Tim Berners-Lee, notes that when he has a new idea he exposes it to others on the Internet by posting his idea to a mailing list or an Internet Relay Chat channel immediately even if such a thought is half-formed. In fact, such an externalization of his thought is a crucial part of his thinking, as the feedback from others over the Internet helps improve his thinking and eventually leads to running code.¹ When there is such a wide divergence in intuitive opinion on such an important question, it is a clear signal that some deft philosophical conceptual work is necessary.

This essay hopes to provide a definite answer to this question of whether the web can extend the mind, and it is not the first to ask the question. Clark and Chalmers point the question out themselves in their seminal work on the Extended Mind: “Is my cognitive state somehow spread across the Internet?” [8]. Clark, Wheeler, and myself then explicitly reposed the argument for the Extended Mind Hypothesis in the context of the Web [13]. Later, Smart analyzed this notion of the Web-Extended Mind with an eye towards the future of evolution of the Web, arguing that “while current forms of the Web may not be particularly suited to the realization of Web extended minds, new forms of user interaction technology as well as new approaches to information representation of the Web do provide promising new opportunities for Web-based forms of cognitive extensions” [22].

What has not been done is a rigorous and nuanced study of under what conditions precisely the Web should count as part of the larger ecological assemblage that constitutes a mind. This will lead us to sharpen arguments not for mere “augmentation” of the mind, but actual cognitive integration of external components such as the Web by seeking by what criteria such technological components could be considered cognitively integrated. There are many cases where external media like the Web are not actually cognitively integrated,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WebSci'13, May 1 – May 5, 2013, Paris, France.
Copyright 2013 ACM 978-1-4503-1889-1 \$10.00.

¹Personal communication with the author.

but simply serve as a way to co-ordinate intelligent problem-solving in terms of what Hutchins calls “distributed cognition.” [16] Yet distributed cognition should not be underestimated, as it is precisely distributed cognition that can solve problems that an individual human may be unable to solve by themselves, and so can serve as a stepping stone to a wider kind of cognitive integration: collective intelligence. Furthermore, we argue that this ability to integrate the mind into media such as the Web is inherently social, insofar as it involves interaction and even tight coupling with other biological human beings. There is a subterranean current in philosophy of the mind based on the under-appreciated work of Vygotsky which posits that it is precisely this externalization into media that is necessary for psychological development in a larger sense, and re-framing the Extended Mind in terms of the Web provides a tantalizing reformulation of Vygotsky’s position that such cognitive integration is actually crucial in developing the mind. Finally, we define collective intelligence as cognitive integration combined with distributed cognition, and give some parting thoughts on its future development.

THE EXTENDED MIND

The question is then: Does cognition - the mind itself - extend into the environment? Two analytic philosophers of the mind, Clark and Chalmers present the example of Otto, a person with short-term memory loss who has to navigate his way to the Museum of Modern Art (MoMA) in New York without the ability to remember its address on 51st Street in Manhattan. Instead of relying on his internal memory, Otto has to rely on his notebook. Given that the notebook performs the same functional role as neural memory, should the notebook be given some of the cognitive credit despite being outside the skin [22]?

The answer seems to be “yes.” Clark and Chalmers also imagine Inga, a woman without such an impairment, who navigates to the MoMA based on her dispositional belief that it is on 51st St. If we agree that Inga’s brain deserves some of the cognitive credit for remembering the address on 51st St., then Clark would argue that it would be an unjustified bias towards neural mechanisms if we would claim that Inga’s mind was responsible for her actions, but not Otto’s mind, simply because Otto’s mind happened to be using a notebook rather than biological memory.

However, one could argue that while Otto’s mind uses the environment, his mind is not properly constituted by extracranial factors such as his notebook. In a well-known critique of the extended mind that defends the notion that (at least currently) only the brain can constitute a mind, Adams and Aizawa claim that Clark commits what they call the “casual-coupling fallacy,” which they illustrate using the following rhetorical question, “Why did the pencil think that $2+2=4$? Clark’s Answer: Because it was coupled to the mathematician” [1]. However, Clark himself can easily respond by simply re-casting the question in terms of coupling between neural elements: Why did the V4 neuron ‘think’ there was a spiral pattern in the stimulus? Because it was coupled to the rest of the monkey! To use an example from Brian Cantwell Smith as regards the difference between a system that is con-

stituted and a system that is simply causally coupled, Smith claims that the test should be an whether or not an organism is destroyed if you “blow up” something to which it is claimed to be constituted [23]. If you remove Inga’s brain, she clearly has ceased in a cognitive sense to be Inga. If you remove Otto’s notebook, does he cease in a real sense in being Otto? At first it seems the answer should be “no.” Upon closer consideration, if Otto’s everyday survival depends on his notebook, would not Otto in a real sense be destroyed if his notebook was stolen?

The radical nature of the Extended Mind Hypothesis should not be under-estimated. The Extended Mind Hypothesis states that cognitive processes, under certain conditions, are constituted by the environment. Thus, an extended mind is a new integrated cognitive whole or “cognitive system” that requires a mutually-reciprocal relationship between the environment and the organism as regards some cognitive function such as memory or attention. This goes far beyond merely claiming the rather obvious point that the mind is embodied (and thus, causally coupled) in a world that also consists of tools needed to solve problems. The claim is that the mind is constituted by such tools, and that far from being bound to the brain or body, the mind is itself an assemblage of such tools. In other words, the mind is a mashup.

The Extended Mind Hypothesis is much more radical than the Embodied Mind Hypothesis, which merely states that “cognitive processes depend very heavily in hitherto unexpected ways, on organismically external props and on the structure of the external environment which cognition takes place” [20]. The key is that an extended mind is constituted by some aspects of the world “outside the skin,” while an embodied and embedded mind merely is deeply causally intertwined with the environment. In other words, if Otto loses his map, he loses mind! However, while many ‘digital natives’ would be distressed if they lost Web access, it would seem a tall-order to claim that losing Web access would literally lead to a loss of their minds. The devil is in the details of relating the Extended Mind Hypothesis to the Web.

THE EXTENDED MIND ON THE WEB

The Extended Mind Hypothesis may seem implausible, but Clark goes to great pains to demonstrate that it only holds under certain conditions. To clarify those conditions is then necessary to determine if and when the Web can be considered to meet these yet undefined conditions and thus can be properly considered constitutive of the mind. What appears to be the main condition is what Clark and Chalmers call the Parity Principle, namely that “If ... a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is (for that time) part of the cognitive process” [8]. Yet the Parity Principle is not meant to be a litmus test for the Extended Mind but merely an “intuition pump” (in the parlance of Daniel Dennett) that is supposed to sharpen our theoretical categories while not by itself serving as definitive analytic test [9]. If we take principles such as the Parity Principle as the actual conditions, various counter-intuitive arguments immediately seem plausible,

ranging from arguments over phenomenology (using a notebook doesn't have the same "feeling" as using neural memory)[10] or to those about implementation details (a notebook does not have the same structure as the brain) [1]. Furthermore, such a misinterpretation of the Extended Mind Hypothesis leads to restrictions on cognitive extension that rule out the Web: namely that digital media such as the Web could not be considered a possible candidate for cognitive integration precisely because digital media have different properties to neurally-implemented mechanisms. Digital memory is a seemingly perfect memory that, unlike our neural memory, neither forgets naturally nor is bounded by its own individual biological structure.

The original Extended Mind Hypothesis is framed in terms of dispositional beliefs, a belief that is stored in memory and recalled when needed, such as the location of the Museum of Modern Art. The original candidate conditions posed by Clark and Chalmers are especially suited for dispositional beliefs, namely that the information is constantly available, directly available without difficulty, automatically endorsed, and consciously endorsed at some point in the past [8]. While the last two points seem reasonable in the case of externalized information capable of forming a dispositional belief, these four aforementioned conditions would rule out many rather sensible cognitive extensions such as one that was not a dispositional belief and thus did not have a sense of being endorsed, such as an extension to attentional capability that allowed better multi-tasking. Lastly, the point of conscious endorsement would seem to be rule out of the Web: after all, we can't consciously endorse every website before we access it over the Web! Or do we stretch the notion of conscious endorsement to the notion of trust in certain sources of information on the Web? In which case, does one trust every website Google indexes, the company itself, or the search engine? It's fairly clear that a substantial revision is needed.

Clark provides four revised conditions to distinguish cognitive extension from mere causal coupling: the conditions of being portable, robust, augmentation, and dovetailed.² Taking a previous example, imagine the case where Otto is not only using a map to go to the Museum of Modern Art, but is being guided by a Web-enabled smartphone that is updated in real-time [13]. If there is a particular subway closure that may prevent him from getting there, real-time notifications can warn him and guide his behavior. Is then the Web part of Otto's extended mind? For Otto does losing his smartphone mean that he may lose some of his mind? Let us inspect the Web in light of Clark's updated criteria:

Portable: The criterion of portability simply means that capabilities that count as part of the extended mind are accessible when needed. This does not mean they are always accessible, but simply accessible under normal conditions; obviously even the brain itself may lose its capabilities over time or be damaged in accidents. While simply accessing the Web with a web-browser on a desktop computer obviously would not count as portable, the Web is clearly increasingly portable

with the advent of portable smartphones with increasing coverage. Otto's portable smartphone is in fact just as portable as his notebook.

Robust: This criterion means that not only are capabilities present when needed, but that they work consistently and within the needed time-frame necessary to solve problems. Thus, a smartphone whose connection failed often or was very slow in delivering information would fail to meet this criteria, as would the use of a search engine that produced irrelevant results. Yet it seems there is some increasing pressure for lower-latency connection to the Web. For example, Google products are famously tested to reduce any delays to less than a 100 milliseconds the time it takes most humans to even notice a delay.³ Also, thanks to machine-learning, it appears results from the Web are increasingly not prone to error. Much like the portability requirement, it seems the Web is increasingly robust. Otto's portable smartphone can be as robust as his notebook, and even moreso as there is no time to be wasted in flipping through pages.

Augmented: This means that the cognitive extension must provide a genuinely new capability or significantly augment an existing capability. For Otto, as he lacks a working short-term memory, the notebook provides a genuinely new and useful capability, as would a smartphone. Contra the Parity Principle, for someone who has a working biological memory, we can still argue digital media such as the Web extend memory and provide memory with distinctly new (digital) characteristics. The speed at which one can retrieve information and communicate globally extends current capabilities (such as pre-Web capabilities of linguistic reading and speech) and this may count as a new capability [5].

Dovetailed: This criterion means that the extension must be evolutionarily co-adapted. This means that a new capability grows and changes with the mind in such a way that it becomes adapted to face both current and anticipated problems, and thus in a sense automatically trusted and endorsed even without necessarily conscious intervention. The Web is clearly adapted to the organism, as personalized algorithms, such as recommender systems and search engines, learn from behavior and are therefore co-adapted to their users.

So, it seems under some circumstances the Web itself can be considered to be part of the mind. Your mind is not extended by just visiting and reading a web-site from a desktop computer. An always-accessible smartphone that uses your geolocation and current set of tasks to deliver you relevant information "just in time" would count as part of your cognitively extended machinery. While the technological infrastructure of the Web is still evolving such that the criteria of cognitive integration are not always fulfilled, and in fact may only be fulfilled for currently small sectors of the population, given the rapid uptake both in users of the Web across the world and the corresponding growth of both reliable infrastructure (as well as increasingly co-adapted tools that provide a host of new capabilities), it seems that the trajectory of the Web is on course with that of cognitive integration.

²Given by Clark in his lecture "Extending the Mind with Cognitive Prosthetics" at La Sorbonne in 2012.

³<http://googleresearch.blogspot.fr/2009/06/speed-matters.html>

Distributed Cognition on the Web

Although the Web may seem to be more and more part of our mind, it seems that our everyday use of the Web is far more mundane: to search for information, to solve problems, and simply keep in touch with friends. Under many conditions, it seems that this ability to use the Web does not fit the characteristics of being a portable, robust, augmented, and dove-tailed part of our mind. Instead, this everyday use of the web is more realistically just another form of technologically-mediated social communication. Yet all may not be as simple as mere communication in the manner of ordinary natural language use, for there are definite structural changes caused in cognition that are accelerated by an increasingly powerful digital medium. Following Hutchins, this case of usage of the web we will call *distributed cognition*, as “groups may have cognitive properties that differ from those individuals who constitute the group” [16]. In other words, even if full cognitive integration is not achieved, there may still be changes to the cognitive properties of the group that make distributed cognitive powers greater than the sum of their individual human parts.

Distributed cognition was first studied by Hutchins in his ground-breaking analysis of how sailors onboard a U.S. Navy ship used various tools, such as astrolabes, and social formations, such as the hierarchical organization of a ship’s commanding officers, to accomplish complex tasks that would otherwise be out of their reach [16]. Although the example may seem limited, the ultimate lessons of Hutchins is how humans can co-ordinate to solve difficult problems like ship navigation that would otherwise seem impossible. Hutchins posits that the essential way humans solve problems in distributed cognition is through the “propagation of representational state across a series of representational media,” where representational state can be considered “the configuration of elements of a medium that can be interpreted as a presentation of something” [16]. This process is actually rather elementary, and the external and representational elements can be something as simple as the human’s fingers. For example, it has been found experimentally that students who were given difficult problems and allowed to gesture with their hands (either spontaneously or encouraged explicitly) performed better than those who did not gesture [18].

Often externalizing a successful solution to a problem in the form of a technological artifact that can be shared amongst others is more useful than embodied externalization that is confined to the particular time and space of the organism. Problems that would otherwise be impossible to solve using the limited biological cognitive capabilities of humans can be recast using elements in a non-biological medium, and then propagated throughout time and space to overcome the limits of human memory and the lifespan of the human body. Hutchins then uses this insight to connect cognition to the transmission of culture across generations. Each culture has over millennia developed their own solutions to be shared: “Solutions to frequently encountered problems are crystallized and saved in the material and conceptual tools of the trade and in the social organization of work” [16]. In this manner, the exact kinds of problems that can be solved are

determined not by the limits of neurons or biology, but by the architecture of the technology.

First, there is the question of the representational medium: the Web’s medium is digital. While a thorough investigation of the properties of the digital is outside the scope and done elsewhere [12], nonetheless it should be stated that one of the reasons for the Web’s success is the choice of a digital representational medium. Earlier visions of Web-like systems, such as Bush’s Memex, imagined to be implemented on analogue mediums like microfiche and never succeeded [6]. Due to the insight of the early Internet pioneer such as Douglas Engelbart, the underlying representational media for externalized memory systems was chosen to be digital. The importance of this insight cannot be underestimated, for as put by Hutchins, some implementation media serve the “externalization of function better than others” [16]. Digital media crucially provides a kind of memory that complements human memory precisely by providing the capability of “flawless copying and perfect reliability” that human memory lacks [12]. Due to its flexibility, digital media are also by nature universalizing media. As the Web allows an open-ended and extensible number of media to be involved, ranging from written text to video to even raw data through an open and extensible number of media types, the kinds of things that can be represented are much more flexible than many other representational medium such as writing and television. Once confined to modality-specific media (pictures and photography, radio, television), due to the invention of digital computers cognitive technology could provide a convergence of all previous media types. The amount of media is growing at an astounding rate, with the growth of digital media outpacing all previously recorded analogue media. Yet as important as the notion of digitality is, the representational medium is simply the substratum upon which representational state are propagated.

It is the architecture of the Web that connects the representational states of various resources together, allowing previously unimaginable amount of digital data stored in the representation medium to be knitted together into a Web. As put by Hutchins, “when the nature of the problem is seen as a coordination amongst person and devices, much of the organization of behavior is removed from the performer and given over the structure of the object or system with which one is co-ordinating” [16]. In this case, the abstract structure of the Web is what allows certain kinds of affordances necessary to distributed cognition.

The key distinguishing characteristic of the Web, in contrast to previous systems, is its claim to be a “universal information space” through which any resource can be identified by a “Uniform Resource Identifier” (URI) [4], originally called “Universal Resource Identifiers.” [2] The Web is actually defined as a space of names, rather than a concrete way of shipping around bits, as the latter defines the Internet rather than the Web. The use of URIs as a naming system is what allows any representational state be linked to any other representational state via hyperlinks between them.

While it would be fashionable to decry as pure hubris the “universal” pretensions of the Web, it is precisely the com-

bination of URIs with a digital medium that are making real the Web's universalizing power. Without digital media, the Web would not be able to actually access representations, but merely refer to them. While the distinction between access and reference has been explained elsewhere [14], the intuition is that with natural language names one in the general case refers to things but can't actually make them appear. While "Tim, come here!" may work in some cases, usually saying "Sir Tim Berners-Lee is the inventor of the Web" does not allow one to instantly access Tim Berners-Lee, especially if one is nowhere near him. Yet URIs provide, under normal network operating conditions, exactly this power over digital objects (not people, at least yet!) due to digitality. Note that many other technical proposals that described naming systems for universal information spaces (Universal Resource Names, Digital Object Identifiers, and the like) have failed to reach widespread usage precisely because they did not actually have built in a mechanism for accessing digital representations. A universal naming system on top of an analogue space of resources without digital representations would not have the explosive growth of the Web. Likewise, one can see the limits of earlier knowledge representation systems, hypertext systems, and even earlier pre-Web Internet systems such as Gopher were the reverse: although they were all built on a digital medium that allowed a space of digital resources, they did not allow an open-ended and extensible naming scheme that could retrieve representations of any digital resource. Earlier systems were considered "walled gardens" where names could only refer to digital resources within the scope of the system's architecture (like Facebook user names). It is the combination of an universal and open-ended naming scheme combined with an open-ended space of digital resources that provides the unique architecture of the Web.

VARIETIES OF DISTRIBUTED COGNITION

Not all systems of distributed cognition are the same. Not only is there the general technological architecture that provides various possible affordances, but there are the particular configurations of humans sharing cognitive resources using these infrastructure. Due to this, despite the abstract possibility offered by the Web that everything can be connected, everything is actually not connected. As science has shown [27], some things are more connected than others; this has been demonstrated on the Web in particular by the emergence of a hyper-connected "giant components" in web-sites, as Google or Yahoo! simply has more outgoing and incoming links than most other websites. The formation of "power-laws"⁴ holds not just for websites but also for human social networks, as some people have thousands of friends on Facebook while the majority of us simply have a hundred or two hundred friends. In any distributed cognitive system, the pattern of connectivity is the first factor that must be taken into account.

The trajectory of the Web is one of ever decreasing latency as well as ever increasing universalizing scope, and so the factor of time should also be taken into account. Not only has the

⁴Within a given network, there is both a minority of densely connected nodes called "the giant component" and a majority of less connected nodes called "the long tail."

amount of digital information grown without bounds on the Web, but there has been a massive decrease in the latency necessary to access digital information due to the spread of high-speed Internet access. Due to mobile smartphones, the current generation of digital natives has always been effectively online, and the psychological and neurological consequences of this are just beginning to be explored. The ontological consequences of this 'digital turn' are still unknown. While once one had to go to a library to physically retrieve a book from a shelf or even put a compact disc into a player in order to listen to music, today it appears as if all of the digital and digitized information produced by humanity can be within the grasp of Internet-connected human within seconds if they are familiar with the retrieval engines such as Google. Today, you no longer even need to search: Thanks to services ranging from Twitter to Foursquare, customized algorithms can instantly and even without interaction bring digital content to that you might want. In this regard, the second factor needed in understanding distributed cognition is temporal latency.

What is necessary to understand distributed cognition is to understand what different cognitive properties result from different patterns of connectivity, latency, and other additional factors such as influence. The results may very well be surprising: Hutchins began simulating distributed cognition via constraint satisfaction networks, where each node was a human and a link some kind of communicative connection [16]. He then tested a number of patterns of connectivity, ranging from highly disconnected to entirely connected, with different speeds of communication. In particular, he was interested in determining, given a set of constraints, how quickly the network would settle on an interpretation that fulfilled the constraint. In order to make his point, rather than complex technological or social constraints, Hutchins focused on simple constraints that could be mathematically formulated (such as "give me a number between one and ten that is even"). He discovered that "more communication is not always in principle better than less. Under some conditions, increasing the richness of communication may result in undesirable properties at the group level," as for highly connected groups with high bandwidth communication, the network "moves towards the interpretation it had moved to in the absence of communication, but now it moves so more quickly...having arrived at that interpretation, they remain there, absolutely unmoved by any amount of evidence from the environment. At high levels of persuasiveness, this system thus manifests a much more extreme form of confirmation bias than any individual alone" [16]. Indeed, it seems there are dangers in building a distributed cognitive system that is "connected" all the time! In detail, the argument against being plugged in all the time is that some degree disconnection and low latency are necessary in order to allow evidence from the world to have sufficient diversity of interpretation of evidence. This is done by "breaking up continuous high bandwidth communication...implemented in social organization, in the interaction of an individual with an external artifact, or through the use of internal mediating structure" [16]. So, what is necessary is what has been called in Web architecture loosely connected components, components of a system that are themselves densely connected but that have only loose connections

to other components. While such experiments studying the relationship of confirmation bias to connectivity and latency in Web networks such as Twitter have only just begun, what is clear is that there is no single homogeneous distributed cognition, but a variety of architectures of different kinds of heterogeneous distributed cognitive systems.

The key question is then what distinguishes the cognitive integration given by the Extended Mind from distributed cognition? Distributed cognition does not necessarily entail cognitive integration, while cognitive integration can be considered a kind of distributed cognition. According to the earlier criteria necessary for cognitive extension, the connections must be portable, robust, accessible, and dove-tailed; many distributed cognitive systems do not fit this rigorous criteria. For example, simply using the Web on a desktop machine occasionally allows you to participate in a distributed cognitive system to solve some kinds of problems, but also the use of hand-written letters delivered by the post office would count a distributed cognitive system of another type. It is only certain precise types of densely interconnected distributed cognitive systems that properly may count as cognitively integrated systems. Thus, it is better to view distributed cognitive systems as a landscape, of which only the most densely interconnected and low-latency systems would count as cognitive integrated systems. Of course, this does bring up the question of whether such a cognitively integrated system would be susceptible to confirmation bias? Of course as such cognitively integrated systems by their very nature need to re-act quickly to new information, just as individual humans had to react quickly to new information about sabre-tooth tigers being nearby in our evolutionary past. It is better to consider cognitively integrated systems to be single systems, which can then themselves take part in a variety of more ad-hoc and loosely connected distributed cognitive systems; these distributed cognitive systems may even over time evolve into cognitively integrated systems.

THE SOCIAL WEB

The focus of distributed cognition on technologically-mediated social interaction as part of cognition is in contrast to most of the work on the Extended Mind Hypothesis, which has in general focused on how an individual biological mind can be extended or constituted cognitively by technology. Clark and Chalmers in their original paper are not in principle against other minds being part of a cognitively integrated extended mind: "Could my mental states be partly constituted by the states of other thinkers? We see no reason why not, in principle" [8]. In the case of the Web, currently the social increasingly predominates. To return to Tim Berners-Lee's excellent example, when Berners-Lee is thinking through an idea, he finds it useful to phrase the concept as best he can in text and put the idea into an Internet Relay Chat channel, where others can discuss and comment on the idea, and so collectively improve and eventually code the idea. The Web seems to be reversing our internalization of thought, with currently more and more thoughts being put on the Web for all to see and comment upon (and otherwise "mashup" and transform). Even if our thoughts are not understood, we nonetheless externalize them on the Web.

Interestingly enough, this phenomenon is also a very common stage in the development of children. Egocentric speech is the tendency for children to speak to themselves when they are accomplishing some task [19]. This phenomenon seems to begin at seven or eight years of age and then disappears afterwards. This particular kind of speech is "externalized" when children encounter problems they seem to need help solving, and it has even been shown that in the presence of solving difficult tasks, the amount of egocentric speech in children increases. This can even be seen in adults when they "talk through a problem" - even to themselves. For example, when discussing how to solve a difficult problem such as determining the most efficient route on a map, even adults routinely will say "Well, if you take this road, you then have to turn left, so perhaps its better to take the other road first..." Why does this externalization help? It can be considered that egocentric speech is helping direct the child's actions, in the same way that a parent's voice directs a child. Following the Extended Mind Hypothesis, the translation of thought to speech at such an early age may be necessary in order for the child to even think complex thoughts, perhaps due to the lack of the necessary biological scaffolding required for thought, via language, to be internalized. Piaget assumed such speech was purely self-directed and individual: "To put it quite simply, we may say that the adult thinks socially, even when he is alone, and that the child under seven thinks egocentrically, even when in the society of others" [19].

A Russian developmental psychologist, Vygotsky, reinterpreted Piaget's findings in a revolutionary manner. He stated that the externalization of speech is not only necessary to direct the child, but that strangely enough egocentric speech is actually social: "In experiments [placing a child] with deaf-mute children or with children speaking a foreign language...the coefficient of egocentric speech [of the child] dropped to zero in the majority of the cases and the rest to one eighth" [26]. To attempt to interpret these findings, Vygotsky theorized that "children who are participants in the collective monologue do believe that they communicate with each other. They believe that their thoughts, even those that are poorly expressed, or unarticulated, belong to all participants" [26]. Between the ages of five and eight egocentric speech becomes internalized as inner speech and thus transforms the capabilities of thought, so that thought can possess various linguistic characteristics such as abstraction and systematicity, properties that Vygotsky theorizes were not present beforehand. Speech is ultimately external before its convergence with and combination with internal thought.

Vygotsky's theory is that functions are first external and social before becoming part of the mind: "Any higher mental function necessarily goes through an external stage in its development because it is initially a social function. This is the centre of the whole problem of internal and external behavior...when we speak of a process, 'external' mean 'social.' Any higher mental function was external because it was social at some point before becoming an internal, truly mental function. It was first a social relation between two people" [25]. Through feedback loops with egocentric speech, the relationship between language and thought is inscribed

neurally. Likewise, the same considerations may very well hold on the Web: tweeting on Twitter and status updates on Facebook can under many circumstances be considered digital forms of egocentric speech. It is hard to argue that some of the all-too-brief short messages on Twitter don't resemble the abbreviated and often disjoint egocentric speech of children!

As humans first encounter their external media, they may begin to use it to help them externalize their own thoughts regardless of the presence of others. For example, Berners-Lee's original pre-Web system, Enquire, was built with the express purpose of helping him organize his own thoughts, not to share them [3]. However, the possible presence of others seems to be a powerful evolutionary motivator, which has led to the most popular sites on the Web like blogs and social networking sites that let people send status updates to each other and comment upon affairs, even if (as is to be expected for most blogs or social networking accounts) their commentary is not being explicitly used in some kind of distributed cognitive problem-solving system. In fact, the most likely case, no-one is even reading their commentary on the Web. Yet over time, this continual egocentric use of the Web provokes a feeling that the Web really is part of their memory, and the kinds of "links" they build to representations on the Web become crucial to their problem-solving behavior, even as individuals. Furthermore, the question is then if egocentric speech allows language to be internalized into a cognitively integrated whole individual, then does the technological mediation of the social web allow larger collective communities to transform into new cognitive systems? The answer should be that the social use of the Web is what changes the entire game. Taking Otto's use of a smartphone, only with the social aspect in place, where other people use the Web, can they update the location and schedule of the Museum of Modern Art in real-time. In order to understand this question, we need to return to the contrast between the collective and the individual, and address the question of collective intelligence.

COLLECTIVE INTELLIGENCE AS COGNITIVE AUTOPOIESIS

The term "collective" in collective intelligence is an empty place-holder, a mark of scientific and philosophical shame. A superficial understanding of collective intelligence simply posits some kind of aggregate in contrast with the individual: the individual versus their larger world, the individual against the crowd, the individual against the totality of existence. Given that theories such as the Extended Mind Hypothesis stretch the boundaries of the individual to what appears to be a breaking point, we only dimly grasp what the outlines of that which ontological category precisely is to come after the individual. Our hypothesis is that collective intelligence is a cognitively integrated distributed cognitive system that involves both other humans and technology, and fulfills our updated criteria for cognitive integration.

Clark and Chalmers provide an initial answer to the troubling Cartesian dichotomy between mind and world given by "I think, therefore I am," namely that a mind can, and always has been, extended into the world and thus technology [8]. Yet Clark and Chalmers fail to extend their argument beyond

the individual in their guiding examples. The logical conclusion of taking their thought experiment seriously points to the hypothesis that there is no individual mind per se, but a process of what Simondon termed "individuation," of the creation of the individual from a continually modifying soft assemblage of bio-technological components [21]. There is a world, therefore I think. Hutchins provides a yet another insights: Groups and whole societies can exhibit genuinely new cognitive capabilities that are embedded in larger networks of social relationships and technology. We are, therefore I am [16]. Vygotsky points out that the process of individuation is always engulfed first and foremost in the social, and so what appears to be the cognitive that Clark "extends" outside the skin originates in the history of the social. We are, therefore I think [25]. In this regard, what the Web provides then is an universalizing space where different social and technological assemblages can form different configurations of connections closer and closer to real-time.

In all the aforementioned contributions, what is missing is the ability to "cut" the world at its joints, and so define what ontological category properly subsumes the individual in the era of the Web. While the conditions of cognitive integration is given by the conditions of Clark, it is not itself defined. The clues to the answer may be found in work by Chilean philosophical biologists Maturana and Varela in their attempt to define a similarly difficult term - "life" itself - in terms of their own neologism, autopoiesis: "living organization is a circular organization which secures the production or maintenance of the components that specify it in such a manner that the product of their functioning is the very same organization that produces them" [17].

As noticed earlier [11], Maturana's concept of autopoiesis features an arbitrary biological boundary and purposefully excludes technical systems. This is due to the common-sense notion that technical machines are not "self-healing" like humans, so that a broken machine does not of its own accord repair itself when damaged. Instead it requires repair by humans. Furthermore, machines in the general sense (outside of research in simulating "artificial life" within very limited simulations) do not reproduce themselves, but require humans to create machine-producing machines (factories) to produce new machines. This would violate Maturana's crucial point that autopoietic systems "through their interactions and transformations continuously regenerate and realize the network of processes (relations) that produced them"[17]. However, there is a crucial way out. First, there is no reason a priori why these interactions have to be biological but can not include technical machines, if those machines were self-healing and reproducing, although this would exclude by definition any techno-biological distributed cognitive system or technically extended mind, as in the case of a Web-extended mind [17]. Stepping back, a little shift can clarify the picture: machines and humans together as an unitary system reproduce yet more machines and humans. Today, many humans would not survive if not for the complex technical medical apparatus, which is itself increasingly intermeshed into the vast amounts of health data available on the Web. Likewise, humans create machines and aid in their repair and reproduc-

tion: not for the good of the machine, but for their own good. This seemingly human good is also the common good of a particular social assemblage (society) of humans and machines.

Another objection to autopoiesis is that it is defined explicitly in reference to a closed system, so that autopoiesis works insofar as its reproduction “constitute it (the machine) as a concrete unity in the space which they (the components) exist by specifying the topological domain of its realization as a network” [16]. In other words, when frogs are wounded, they heal to be more-or-less (depending on the severity of the wound) a frog, and frog parents give birth to baby frogs rather than a whole new kind of creature. Somehow, an autopoietic system must contain its own blue-print, which Maturana would presume is done in the DNA. Yet this picture of autopoietic systems as closed is correct only for the lifespan of an individual in the species, not for the species as a whole, otherwise the theory of autopoiesis would reject Darwinian evolution. Second, even Maturana admits that biological species have to interact with the outside environment; frogs eat gadflies and humans use the Web to get directions to Centre Pompidou. This is taken into account in autopoietic theory by what they term the “structural coupling” that defines an organisms interaction with the environment as a “history of recurrent interactions leading to the structural congruence between two (or more) systems” [16]. Instead of a closed systems perturbing each other for their mutual autopoiesis, envisage a “little shift” in perspective that views these two individual systems as a single integrated system maintaining autopoiesis jointly.

The question is what makes such an autopoietic system cognitive? Obviously many autopoietic systems are non-cognitive: the reproduction and maintenance of a single-celled bacterium would count as autopoietic, but there seems intuitively little to do with cognitive factors such as memory, attention, or language in a bacteria. An autopoietic system would only count as cognitive if the self-sustaining feedback loops involves language, memory, and attention. Clark calls this “cognitive self-stimulation” [7], which occurs when a system is causally responsible for producing structures and events which are then recycled as inputs to the very cognitive system itself and thus sustains sophisticated (brain-body-environment) loops of exploitation and mutual co-ordination that can help solve particular problems. The boundary between an input (perception) to a system and the output (action) of a system melts and the interface becomes part of the system. We have already encountered a relatively simple example of this in the form of egocentric speech as explored by Vygotsky, where the child’s speech serves as the cognitive self-stimulation necessary to solve some problems. More complex examples can clearly involve the Web, such as the use of Twitter to co-ordinate a distributed cognitive organization involving many humans or, in the case of full-scale cognitive integration, the use of the Web-enabled smartphone by Otto to guide him to various locations. We find that examples like cognitive self-stimulation may let us consider autopoietic systems to not be eternal and unchanging biological species-beings, but open bio-technological assemblages.

If the individual can be defined via autopoiesis, and to maintain its autopoiesis the individual can increasingly incorporate environmental components, then the individual is no longer a static, closed system, but an open and dynamic system capable of assimilating and decoupling from various components as it goes in and out of autopoiesis - and this includes digital representations on the Web. Crucially, this integration also may include other humans. Then we can define collective intelligence as a form of autopoiesis that goes beyond the skin of a single individual as to create a self-reproducing and modifying unified intelligence that includes connecting multiple individuals via technology. As the Web is one of the few technologies that satisfies the criteria of Clark’s Extended Mind (portable, robust, co-adaptive, augmented) the Web is a natural medium for collective intelligence.

CONCLUSIONS

Intelligence has always been collective. The advent of the Web, an universal digital medium that contains within it the latent possibility of connecting all of humanity and the world as and when necessary, arrives just in time to disrupt our previously stable Enlightenment ontology of individual. However, a new scientifically-grounded definition of collective intelligence is just one task amongst many, for at the present moment we lack a thorough understanding of the transformation of humanity at the hands of digital media, and such an understanding goes far beyond cognitive science. We have argued that certain conditions (those of being portable, robust, accessible, and dove-tailed) distinguish cognitive integration within the larger landscape of distributed cognition. We further argued how even linguistic phenomena can play a possibly technological mediated role in communication. Lastly, we defined collective intelligence in terms of techno-social self-organization and reproduction.

The stakes of answering the question of ‘does the web extend the mind?’ are very high. As discussed by Stiegler, the ability of our minds to use and integrate what appears to be external media not only underpins much of intelligence but also our understanding of wider social and ethical questions [24]. While at first it may seem that programmers like Berners-Lee are in the minority in terms of feeling that the Web is part of their mind, it may very well be the case that the intuitions of a Web-savvy minority of humanity will become the intuitions of the majority of humanity within a generation.

Collective intelligence requires that an individual mature to reproduce not only biologically, but their larger techno-social cognitive niche. This successful reproduction (which requires both connecting and disconnecting to others) is the key of long circuits necessary for Stiegler’s collective individuation [24]. Yet does this Web-based communication lead to collective individuation? Perhaps not: “When users are considered as social atoms superimposed onto a technological network, the spontaneity and innovation within their possible collective intelligence is deformed by the control of the networks, driven as it is by intensive marketing and consumerism aimed at individuals rather than the development of the potential of the group” [15].

Given the financial crisis and catastrophic climate change,

never has the very reproduction of our world been so threatened, and yet never has it been greeted with such indifference. This reason is that the problems facing us seem literally impossible to even imagine: the magnitude of the task for coming generations dwarfs the cognitive grasp of any single individual and our pre-Web social institutions. Currently our technologies, from social networking sites like Facebook to search engines like Google, are inadequate, founded as they are on outmoded ontologies of the socially atomized individual and disembodied information.

A new philosophically-informed engineering must arise that can create the kinds of cognitive niches that foster collective individuation and harness collective intelligence to solve problems on a global scale. Even more importantly, our current ethics, based on the notion of the self-interested individual, is clearly at fault for the current crisis; a new ethics must be re-invented that does justice to the dense “intertwingling” between our sense of self, our technology, and our social relationships. It is not enough to define collective intelligence, or to learn how to (re)produce it, but we must also care for our intelligence and our world.

REFERENCES

- Adams, F., and Aizawa, K. The bounds of cognition. *Philosophical Psychology* 14 (2001), 43–64.
- Berners-Lee, T. IETF RFC 1630 Universal Resource Identifier (URI), 1994. <http://www.ietf.org/rfc/rfc1630.txt> (Last accessed on May 3rd 2008).
- Berners-Lee, T. *Weaving the Web*. Texere Publishing, London, 2000.
- Berners-Lee, T., Fielding, R., and Masinter, L. IETF RFC 3986 Uniform Resource Identifier (URI): Generic Syntax, January 2005. <http://www.ietf.org/rfc/rfc3986.txt>(Last accessed on April 2th 2008).
- Berners-Lee, T., and Halpin, H. Defend the Web. In *Digital Enlightenment Yearbook*, J. Bus, M. Crompton, M. Hildebrandt, and G. Metakides, Eds. IOS Press, Berlin, 2012, 3–7.
- Bush, V. As we may think. *Atlantic Monthly* 1, 176 (July 1945), 101–108.
- Clark, A. *Supersizing the Mind*. Oxford University Press, Oxford, 2010.
- Clark, A., and Chalmers, D. The Extended Mind. *Analysis* 58, 1 (1998), 7–19.
- Dennett, D. *Consciousness Explained*. The Penguin Press, Allen Lane, 1991.
- Dror, I., and Harnad, S. Offloading cognition onto cognitive technology. In *Cognition Distributed: How Cognitive Technology Extends Our Minds*, I. Dror and S. H. Harnad, Eds. John Benjamins, 2008, 1–23.
- Halpin, H. Foundations of a philosophy of collective intelligence. In *Proceedings of Convention for the Society for the Study of Artificial Intelligence and Simulation of Behavior* (2008).
- Halpin, H. Becoming Digital: Reconciling theories of digital representation and embodiment. In *Philosophy and Theory of Artificial Intelligence 2013*, V. Mueller, Ed. Springer, Berlin, 2013, 199–213.
- Halpin, H., Clark, A., and Wheeler, M. Towards a Philosophy of the Web: Representation, Enaction, Collective intelligence. In *Proceedings of the Web Science Conference: Extending the Frontiers of Society On-Line (WebSci 2010)* (Raleigh, North Carolina, USA, 2010), 1–5.
- Hayes, P., and Halpin, H. In defense of ambiguity. *International Journal of Semantic Web and Information Systems* 4, 3 (2008).
- Hui, Y., and Halpin, H. Collective Individuation: The Future of the Social Web. In *The Unlike Us Reader*, G. Lovink and M. Rasch, Eds. Institute for Network Cultures, Amsterdam, 2013, 103–116.
- Hutchins, E. *Cognition in the Wild*. MIT Press, Cambridge, Massachusetts, 1995.
- Maturana, H., and Varela, F. *Autopoiesis and Cognition: The Realization of the Living*. D. Reidel Publishing, Dordrecht, 1973.
- McNeill, D. *Gesture and Thought*. University of Chicago Press, Chicago, 2005.
- Piaget, J. *The language and thought of the child*. Routledge, London, 1959.
- Rupert, R. Challenges to the Hypothesis of Extended Cognition. *Journal of Philosophy* 101 (2004), 389–428.
- Simondon, G. *L'individuation psychique et collective*. Aubier, Paris, 1989.
- Smart, P. The Web-Extended Mind. *Metaphilosophy* 43 (2012), 446–463.
- Smith, B. C. *On the Origin of Objects*. MIT Press, Cambridge, Massachusetts, 1996.
- Stiegler, B. *Technics and Time: Volume 1*. Stanford University Press, Palo Alto, 1998. Translated by Richard Beardsworth and George Collins.
- Vygotsky, L. *Mind in society: The development of higher psychological processes*. Harvard University Press, Cambridge, 1978.
- Vygotsky, L. *Thought and Language*. MIT Press, Cambridge, 1986.
- Watts, D., and Strogatz, S. Collective dynamics of 'small-world' networks. *Nature* 393, 6684 (1998), 440–442.