

Smart, P. R. (2018) 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.

14

Emerging Digital Technologies Implications for Extended Conceptions of Cognition and Knowledge

Paul Smart

Introduction

Digital technologies play an increasingly important role in shaping the profile of human thought and action. In the few short decades since its invention, the World Wide Web, for example, has transformed the way we shop, date, socialize, and undertake scientific endeavors.¹ We are also witnessing an unprecedented rate of technological innovation and change, driven, at least in part, by exponential rates of growth in computing power and performance.² The technological landscape is thus a highly dynamic one—new technologies are being introduced all the time, and the rate of change looks set to continue unabated. In view of all this, it is natural to wonder about the effects of new technology on both ourselves and the societies in which we live.

When it comes to the social impacts of new technology, an interesting debate has emerged regarding the extent to which technologies are, on occasion, incorporated into the fabric of society (e.g., Smart and Shadbolt 2014). Consider, for example, the case of the Web. The first uses of the Web did not yield a sudden change in the organization and operation of society; rather, the Web served as something of a technological add-on to an existing set of social practices and conventions. At this point, the project to create

¹ I am indebted to Spyridon Orestis Palermos for his feedback on an earlier draft of this paper. I would also like to thank the attendees of the 1st International Conference on Extended Knowledge in Edinburgh, Scotland for their insights and comments. 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.

² One insight into the shape of the future technological landscape is provided by a number of empirically derived laws. These include Moore's Law (which predicts future advances in computing capacity), Kryder's Law (which predicts future increases in digital storage capacity), Cooper's Law (which predicts future increases in wireless communications capacity), and Nielsen's Law (which predicts future improvements in Internet connection speeds). In general, all of these laws predict exponential rates of growth and performance over specific timeframes.

the Web could have been abandoned and society (as it existed at that time) would have been largely unaffected. Fast forward thirty years, however, and it seems much harder to decouple the Web from society. Indeed, the Web now appears to play a role in society akin to that played by the systems that transport food, water, and energy. Like these relatively long-standing systems, the Web now constitutes a form of critical infrastructure for society: without it, society (at least in its current form) could not survive. What seems to have happened is that, over time, the Web has become increasingly intertwined with the structures and processes that make our society what it is. So deep is this level of entanglement that it now seems difficult to separate our notions of what society is from the various forms of technological scaffolding that the Web provides. It is almost as if the Web had become incorporated into the very fabric of society, making it difficult to discern where the realm of the social stops and the realm of the technological begins. Social and technological processes have, it seems, become inextricably linked, and both have perhaps been transformed as a result of the socio-technological merger.

Issues of entanglement, transformation, and incorporation thus lie at the heart of recent attempts to understand the impact of new technologies on our social world. These issues, however, also lie at the heart of another debate concerning the impact of new technologies on our individual cognitive and epistemic capabilities. Of particular interest is the idea that bio-external resources can, on occasion, participate in the mechanistic realization of human mental states and processes. The idea, in essence, is that the machinery of the human mind can extend beyond the biological borders of the body to include a variety of non-biological elements (Clark 2008). On the back of such claims, we can ask whether emerging digital technologies might contribute to the emergence of what are called “extended cognitive systems” or “extended minds.” In other words, we can ask whether the properties of emerging digital technologies are consistent with the sorts of criteria that are required in order for cognitively potent forms of bio-technological merger to occur. Advancing our understanding in this area is a key aim of the present chapter. The present chapter also aims to examine the nature of the relationship between emerging digital technologies and philosophical conceptions of extended knowledge (Palermos and Pritchard 2013; Palermos 2011; 2014a). I will suggest that by reflecting on technologically advanced forms of cognitive extension we are able to bring a number of important epistemological issues into sharper focus.

1. Extended Cognition

1.1 Cognitive extension

The notion of cognitive extension captures the idea that the causally active physical vehicles of cognition can, at least in principle, extend beyond the biological borders of the individual and encompass a range of extra-organismic resources (Clark 2008; Clark and Chalmers 1998). This idea is sometimes presented as a thesis about the

explanatory kinds of interest to cognitive science (in which case it is referred to as the hypothesis of extended cognition), and sometimes it is presented as a thesis about mentalistic folk categories, such as states of belief (in this case, it is commonly referred to as the extended mind thesis).

As a thesis about mentalistic folk categories, the notion of cognitive extension is used to motivate the claim that extra-organismic resources, such as notebooks, iPhones, and so on, can, under certain circumstances, form part of the mechanistic supervenience base for common mental states, such as states of dispositional belief. This idea was first outlined by Clark and Chalmers (1998) in their seminal paper, titled “The Extended Mind.” As part of that paper, Clark and Chalmers asked us to imagine two individuals, Inga and Otto, both of whom are situated in New York City. Inga is a normal human agent with all the usual cognitive competences. Otto, however, suffers from a mild form of dementia and is thus impaired when it comes to certain acts of information storage and recall. As a means of coping with this impairment, Otto relies on a conventional notebook that he uses to store important pieces of information. Otto is so reliant on the notebook and so accustomed to using it that he carries the notebook with him wherever he goes and accesses the notebook fluently and automatically whenever he needs to do so. Having thus set the scene, Clark and Chalmers (1998) asked us to imagine a case where both Otto and Inga wish to visit the Museum of Modern Art (MoMA). Inga thinks for a moment, recalls that the museum is on 53rd St., and she then walks to the museum. It is clear that in making this episode of behavior intelligible (or psychologically transparent) to us Inga must have *desired* to go to the museum, and it is clear that she walked to 53rd St. because that is where she *believed* the museum was located. Obviously, Inga did not believe that the museum was on 53rd St. in an occurrent sense (i.e., she has not spent her entire life consciously thinking about the museum’s location); rather, she entertained the belief in a dispositional sense. Inga’s belief, like perhaps many of her beliefs, was sitting in memory, waiting to be accessed as and when needed.

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

Clark and Chalmers (1998) argue that the Otto thought experiment establishes the case for a form of externalism about Otto’s states of dispositional believing. The notebook, they suggest, plays a role that is functionally akin to the role played by Inga’s onboard bio-memory. If this is indeed the case, then it seems to make sense to see the

notebook as part of the material supervenience base for some of Otto's mental states, specifically his states of dispositional belief (such as those involving museum-related facts). The main point of the argument is thus to establish a (potential) role for non-biological artifacts in realizing certain kinds of mental states and processes. If, as Clark and Chalmers (1998) argue, the functional contribution of a bio-external resource is the same as that provided by some inner resource, then it seems unreasonable to restrict the material mechanisms of the mind to the inner, neural realm. Rather it seems possible, at least in principle, for the machinery of the human mind to occasionally escape its cranial confines and extend out into the world.

1.2 Criteria for cognitive extension

One of the issues that has dominated philosophical debates about cognitive extension is the notorious problem of “cognitive bloat.” Cognitive bloat refers to an unwelcome expansion in the kinds of resources that feature as the causally active physical vehicles of mental states and processes. As Clark (1997, 217) rightly notes, “[t]here would be little value in an analysis that credited me with knowing all the facts in the Encyclopaedia Britannica just because I paid the monthly installments and found space for it in my garage.” In this case, claims about the encyclopedia forming part of the realization base for one's states of dispositional belief look to be inappropriate.

In order to guard against the problem of cognitive bloat, a number of criteria have been proposed to help distinguish genuine cases of cognitive extension from those of a more ersatz variety. Some of the more popular criteria that have been discussed in the philosophical literature include the following:

- **Availability:** the external resource should be “reliably available and typically invoked” (Clark 2010b, 46).
- **Accessibility:** the “information contained in the resource should be easily accessible as and when required” (Clark 2010b, 46).
- **Trust:** the information retrieved from the resource should be “more-or-less automatically endorsed. It should not usually be subject to critical scrutiny (unlike the opinions of other people, for example). It should be deemed about as trustworthy as something retrieved clearly from biological memory” (Clark 2010b, 46).
- **Personalization:** the resource should be “customised to an agent's individual usage and at the same time the agent's own cognitive routines and predispositions... [should be] altered to incorporate the resulting personalised artefact” (Clowes 2015, 274).
- **Continuous Reciprocal Causation:** “continuous mutual interactions between the organismic agent and the artefact... [are] both necessary and sufficient for cognitive extension” (Palermos 2014b, 26).

This is by no means an exhaustive list of criteria. Neither is it a list that everyone would accept as valid. Nevertheless, the list does provide us with a rough guide as to the sorts of conditions that might need to be met in order for a non-biological resource

to count as part of the material fabric of a cognitive system. With this in mind, my aim in what follows will be to assess how a range of emerging digital technologies fare with respect to these criteria. Specifically, I will be concerned with the way in which the properties of emerging digital technologies might enable them to support current or near-future forms of cognitive extension.

1.3 *Web-extended minds?*

Cognitive extension has emerged as an important focus area for research within the philosophy of mind, cognitive science, and, more recently, epistemology. Within computer science, however, the topic of cognitive extension has been the focus of less attention. This is unfortunate, since digital technologies play an increasingly important role in shaping the profile of human thought and action, and it is important that we understand the role of such technologies in enhancing (or perhaps undermining) our cognitive capabilities at both the individual and collective levels.

An issue of particular importance concerns the applicability of extended mind theorizing to the World Wide Web. Given that the Web now serves as part of the informational backdrop for a range of cognitive, social, and scientific endeavors, it is natural to wonder to what extent the notion of cognitive extension is applicable to Web technology. Can the Web form part of the machinery of our minds, yielding what has been referred to as Web-extended minds (Smart 2012)? If so, how does this affect our epistemic and cognitive profiles? Could the Web herald a new era in which our body of knowledge and beliefs is limited only by whatever information the Web makes available? Could we, in other words, come to enjoy new forms of epistemic omniscience simply by virtue of the kinds of informational contact we have with the online digital environment?

One finds a positive response to these sorts of questions in the work of a number of philosophers who have sought to apply active externalism to the Web. Ludwig (2015), for example, argues that we should embrace the possibility of Web-extended minds and accept the implications this has for our epistemic standing, which he describes as “a staggering explosion of dispositional beliefs and knowledge” (356). Similar views are expressed by Bjerring and Pedersen (2014), who argue that the Web enables us to enjoy restricted forms of epistemic omniscience—a state of affairs in which we have more-or-less “complete knowledge about a particular, fairly specific subject matter” (25).

Are such views correct? Does the nature of our contact with the Web imply an explosion in our body of (dispositional) beliefs and knowledge? One way of answering this question is to assess whether the nature of our current contact with the Web satisfies the sort of criteria that have been proposed to evaluate putative cases of cognitive extension (e.g., the aforementioned criteria of availability, accessibility, trust, and so on). Recall that the motivation for specifying such criteria in the first place was to avoid the kinds of rampant cognitive expansionism that fuels the claims of both Ludwig (2015) and, to some extent, Bjerring and Pedersen (2014). Unfortunately, attempts to evaluate the extent to which Web technologies satisfy the criteria for cognitive extension

have yielded a range of pessimistic, or at least somewhat conservative, conclusions. Clark and Chalmers (1998), for example, appear to dismiss the possibility of Web-extended minds on the grounds that the kind of contact we have with the Internet is likely to violate many of the criteria for cognitive extension—at least the ones pertaining to availability, accessibility, and trust. In more recent work, Clark (2010b) suggests that mobile access to Google should not be viewed as a form of cognitive extension on the grounds that it fails to meet the trust criterion. In particular, he suggests that we don't trust Web-accessible information to the same extent that we trust information retrieved from bio-memory.

My own contribution to this debate some years ago (Smart 2012) was somewhat less pessimistic in tone. I too, however, voiced concerns about the extent to which the nature of our informational contact with the Web could satisfy the criteria for cognitive extension. One such concern relates to the accessibility of online information. Thus, even in situations where a user manages to locate and access a relevant source of information (e.g., a particular Web page) they are still confronted with the rather onerous task of surveying the retrieved 'document' for relevant information content. In most cases, this requires the user to scroll through the Web page and process large amounts of largely irrelevant content in order to identify the small amount of information that is actually required for the completion of a cognitive task. Such forms of interaction typify our attempts to retrieve information from a range of epistemically potent online resources, such as Wikipedia. However, the situation is often worse than this. In many cases, our attempts at information retrieval are interrupted, and sometimes irrevocably thwarted, by all manner of popups and overlay screens that demand some sort of task-irrelevant user response. And, if that wasn't enough, many Web pages are adorned with a variety of flashy ads, banners, and other gewgaws that at best serve as minor distractions and at worst curtail our attempts at information retrieval. All of this speaks against the kind of easy access to information that is mandated by the accessibility criterion.

In responding to these concerns, Ludwig (2015) suggests that accessibility worries are somewhat overblown and that we *can* rely on certain bodies of online information to be accessible in the way demanded by the accessibility criterion. He usefully points out that whenever we want to retrieve (or, in the context of a Web-extended mind, recall) specific facts, such as the fact that Charles Darwin was born in 1809, we do not need to spend a lot of time scrolling through the Wikipedia article on Charles Darwin. This is because the relevant nugget of information is conveniently situated at the top of the article and is pretty much the first fact we encounter. Darwin's birth date, as a piece of factual information, therefore seems highly accessible, and this, Ludwig (2015) suggests, makes it an appropriate candidate for inclusion in our body of dispositional beliefs.

While I am largely sympathetic to the point that Ludwig is making here, I suggest that tying accessibility to the structural layout of a dynamically generated information resource is a risky business. Particular facts may, by virtue of their spatial position, be highly accessible at one particular point in time, but then become far less accessible as

the content is revised and restructured as a result of subsequent (communal) editing actions. In addition, I would suggest that by tying notions of accessibility to issues of page layout and design, the range of facts that could come to form part of our extended body of beliefs is in fact far more limited than is implied by Ludwig's (2015) rather upbeat analysis. It is unclear, for example, whether we should see the number of facts that are highly accessible as legitimating claims that we are poised to experience a "staggering explosion" in our body of dispositional beliefs and knowledge.

Despite the rather cautious tone that was expressed in my original treatment of Web-based forms of cognitive extension (Smart 2012), nothing in that earlier paper excludes the possibility of cognitively potent forms of Web-based bio-technological merger. My point was simply that a particular form of Web-based interaction—one characterized by the use of conventional Web browser technologies and document-centered approaches to information representation and display—is unlikely to serve as a particularly fertile substrate for the emergence of Web-extended minds. We should not, however, be misled into thinking that our traditional modes of accessing the Web tell us everything about the nature of our future contact with the online environment. Crucially, the Web is a system that is in a state of more-or-less constant technological flux. Although it might be useful to think of the term "World Wide Web" as denoting a particular kind of system, the term is, in fact, likely to subsume a highly disparate array of systems, each varying with respect to the kind of interactivity and connectivity they provide to their relevant user base. The 'Web' is, as such, something of a protean beast when it comes to the kinds of interactive opportunities it provides. We would do well to remember this when we evaluate blanket statements to the effect that the 'Web' is (e.g.,) extending our minds (Smart 2012), undermining our cognitive capabilities (Carr 2010), or enhancing our epistemic standing (Bjerring and Pedersen 2014; Ludwig 2015).

1.4 Emerging technologies

In this section, I attempt to provide an overview of some of the technologies that are likely to impact the nature of our informational contact with the Web in coming years. In addition, I will seek to highlight some of the properties of emerging digital technologies that I believe speak directly to the notion of cognitive extension and the possibility of Web-extended minds.

1.4.1 LINKED DATA AND THE SEMANTIC WEB

When most of us think about the Web the thing that comes to mind is a set of Web pages that contain various forms of content—text, images, embedded video, and so on. Some of that content is clearly associated with hyperlinks, and these provide easy access to content that is contained in other Web pages. The Web pages themselves are identified by Uniform Resource Locators (URLs) that specify the 'location' of the page on the Web. It is this URL that enables hyperlinks to 'point to' specific Web-accessible resources. This admittedly crude characterization encourages us to think of the Web as

a globally distributed network of documents connected by hyperlinks. Let us call this the document-centered Web, or the Web of Documents.

In addition to the document-centered Web, another kind of Web has been the focus of recent research attention. This is the data-centered Web, or the Web of Data (Bizer 2009; Heath and Bizer 2011). The basic idea, here, is that instead of viewing the Web as a network of documents, we can also see it as a network of data, with each node in the network corresponding to a body of machine-readable data that is identified by a Uniform Resource Identifier (URI). The idea, in essence, is to view the Web as a globally distributed repository of inter-linked data assets. This idea lies at the heart of an important body of work in computer science that deals with issues of data access and exploitation in Web-based environments. A key aim of research in this area is to understand the way in which the Web serves as the technological substrate for informational ecologies that support a rich array of social, scientific, and political activities.

The precursor to research on linked data is an initiative that goes by the name of the Semantic Web (Berners-Lee et al. 2001). The goal of the Semantic Web initiative is to provide a means of representing information on the Web in a form that is both accessible to machines (i.e., machine-readable) and also supportive of particular kinds of semantically constrained reasoning. Although the Semantic Web and linked data initiatives are sometimes viewed as separate areas of research, it is important to note that both initiatives rely on a common set of technologies that are used to represent, access, and query online resources. In addition, a major focus of attention within the Semantic Web community is the development of computational ontologies³ that seek to represent human knowledge in a semantically enriched format. These ontologies can be used to inform the ‘interpretation’ of data in the context of the Web of Data. An ontology can, in other words, serve as something of a semantic overlay that makes a body of data more amenable to machine-based forms of reasoning and inference.

Why might these developments be of any interest or relevance to philosophers concerned with Web-based forms of cognitive extension and extended knowledge? There are, in fact, a number of points of interest here. Perhaps the most important thing to note is that once we start to talk about the Web as a globally distributed database, we are in a position to address at least some of the accessibility constraints that, it was suggested, bedeviled the information retrieval efforts of human subjects in the context of the document-centered Web (see Smart 2012). In particular, by representing online content (such as that provided by Wikipedia) in a linked data format, e.g., DBpedia (Bizer et al. 2009), we have the opportunity to interact more directly with specific items of information (e.g., facts) than would otherwise be the case. By making Wikipedia available as a form of online structured database, we thus open the door to

³ A computational ontology, in this case, refers to a description of the concepts and relationships that exist in some domain of discourse. A computational ontology typically seeks to represent concepts in a semantically enriched form in order to support various sorts of logical reasoning. Note that the notion of ontology in computer science is distinct from the notion of ontology in philosophy.

a range of novel forms of information retrieval and navigation. Consider, for example, a system designed to support the navigation of large structured data repositories: the mSpace Navigator (schraefel et al. 2005). mSpace uses a combination of linked data assets, semantic queries, and intuitive user interaction protocols in order to support fast, fluent, and easy access to specific items of information that exist as part of a much larger corpus of inter-linked data.

A further way in which linked data is relevant to debates concerning the possibility of Web-extended minds derives from the way in which data-centric approaches to information representation enable us to separate issues of data content from issues of data presentation. Thus, rather than restricting ourselves to the presentational layouts adopted by conventional Web pages, we can begin to consider the ways in which data can be used to drive the creation of a potentially unlimited array of informational artifacts, each of which can be adapted to suit the demands of real-world problem-solving. We already encounter examples of this form of presentational flexibility in the context of dynamically generated Web pages, where the content and layout of the target page varies with respect to the device (e.g., laptop versus smartphone) that is used to access it. Linked data assets can obviously support this kind of device-related flexibility. In addition, however, linked data assets are largely agnostic with respect to the kind of software application that is used to present information content. Such forms of application-neutrality and device-independence are of crucial importance when we consider 1) the broad array of Web-enabled devices that are starting to emerge, and 2) the many and varied task contexts in which a particular online data asset might be called into cognitive service.

1.4.2 CLOUD COMPUTING

Cloud computing is one of the elements of the computer science lexicon that has become something of a household name in recent years. In essence the term “cloud computing” refers to the practice of transitioning data storage and computational services to a remote, networked computer, one that is often managed and maintained by a third party, such as Amazon, Microsoft or Google. Once situated within the “cloud,” data and services are accessible to users in a manner that is typically independent of physical location or mode of access (e.g., the use of a particular kind of device).

It should be relatively clear that the main implication of cloud computing vis-à-vis issues of cognitive extension is that it (potentially) enhances both the availability and accessibility of information. Cloud computing therefore seems to affect the functional poise of online information in a way that is broadly compatible with the possibility of Web-extended minds. There is, however, something of additional interest here. Note that the typical wording of the availability criterion leads us to focus our attention on a particular resource. It is thus Otto’s *notebook* that must be reliably available and typically invoked. The reason for focusing on the notebook (*qua* resource) is that, in the Otto case, the notebook and its constituent informational encodings are inextricably linked: wherever the notebook goes, the informational encodings are sure to

follow. In the case of cloud computing, however, we encounter a rather deliberate attempt to separate the informational encodings from the actual resources (i.e., the technological devices) that are used to access those encodings. In the case of cloud computing, the availability of online information is thus enhanced, not as a result of a particular device (a smartphone let's say) being more-or-less constantly associated with an agent, but rather as a result of the way in which a target body of information can be accessed by a plethora of different devices. It is thus the availability of the *information* that counts here, as opposed to the availability of the *device* that is used to access the information. This doesn't mean, of course, that issues of device availability are entirely irrelevant: absent any kind of Web-accessible device and the target information may as well be stuck on a computer hard drive on the other side of the planet! The point, for present purposes, is simply that cloud computing enhances the availability of information by loosening the bonds that otherwise restrict data access to particular devices and physical locations.

The potential relevance of cloud computing to issues of cognitive extension has been recognized by a number of theorists, most notably by Clowes (2015). In an attempt to better understand the extent to which cloud computing technologies impact our cognitive and epistemic capabilities, Clowes (2015) identifies a range of features that distinguish online, cloud-based forms of information storage from those that are typically encountered in offline contexts. One of these features is "autonomy," which refers to the way in which cloud-based information is amenable to various forms of automated processing. The kinds of processing Clowes (2015) has in mind here are those that work to enhance the functional poise of online content, e.g., tagging, indexing, and restructuring. In addition to this mix of information-processing routines, we can, I suggest, also add the kinds of processing that are afforded by linked data formats and semantically enriched representations. Thus, in situations where cloud-based data storage assumes the form of a linked data repository, and we also have at our disposal one or more computational ontologies that provide a semantic overlay for the repository contents, we can easily imagine online data being subject to forms of processing that include, e.g., checks for logical consistency and coherence, the removal of redundant or outdated data, the assertion of logical entailments based on data semantics, and the automatic updating of repository contents from external (linked) data sources.

A final point to note is that while the term "cloud computing" is typically applied to organizational settings, where multiple users access a common body of data and services, there is no reason why individual users cannot exploit cloud-based technology for the purposes of storing and accessing bodies of personal data. We thus encounter the notions of "personal data stores" and "private clouds," both of which speak to the current interest in enabling individuals to create and manage so-called "personal information ecosystems" (Van Kleek et al. 2012). A particularly important example of this sort of system is SemperWiki (Oren 2005), which combines an online personal data store with semantically enriched content and a Wiki-like editing capability. These

sorts of systems are emblematic of the way in which the Web can be used to provide access to highly personalized bodies of information and data.

1.4.3 WEARABLE TECHNOLOGY

Wearable technologies epitomize the current trend toward greater mobility and portability in the design of digital technologies. A bewildering array of such technologies is now available, including smart watches, EEG headsets, activity trackers, smart clothes, digital jewelry, and so on. Given that these technologies are, by definition, designed to be worn by a user, it should be clear that wearable technologies are relevant to issues concerning the availability and, to some extent, the accessibility of information. In some cases, it is even possible to think of wearable technologies as satisfying the criteria of availability and accessibility to a greater extent than, e.g., Otto's notebook. Otto's notebook, for example, is something that he needs to carry around with him, and it is therefore liable to interfere with many of his other daily activities. Wearable devices, in contrast, are much less obtrusive and can easily fall outside the scope of our conscious awareness, pretty much as our clothes and conventional wristwatches are hardly ever the focus of conscious attention and active scrutiny.

It is also important to note that the progressive encroachment of technologies on our physical selves provides a range of new opportunities for technology to support the end user. In particular, the sensory capabilities of wearable devices (as well as other portable technologies) enable computational processing to be sensitized to various features of the context in which such processing occurs. A trivial example is provided by location-aware devices. Thus, when I am located in New York, and I use my co-located iPhone to look up nearby museums, I do not need to specify that I am in New York. The device, by virtue of its GPS capabilities already 'knows' my location and can thus contextualize the search effort, yielding a set of geographically filtered search results that are likely to be highly germane to my current interests and concerns. Such forms of contextualized processing are likely to become increasingly sophisticated as new sensor-enabled wearable devices begin to emerge. Future capabilities could include the contextualization of information retrieval processes to a number of social, behavioral, and physiological cues. Such sensitivity may even extend into the realm of our epistemic activities. It is known, for example, that certain kinds of epistemic feeling, such as the feeling of knowing or the feeling of difficulty (Michaelian and Arango-Muñoz 2014), are associated with particular kinds of physiological response. The tension of the corrugator muscle, for example, appears to correlate with the subjective experience of knowing (Koriat and Nussinson 2009). By being able to detect such physiological responses, future technologies could adaptively modify their modes of operation to support human end users with regard to a variety of epistemic activities.

1.4.4 AUGMENTED REALITY AND VIRTUAL WORLDS

Technologies that target the realm of augmented, virtual, and mixed reality have been the focus of considerable public interest and media attention in recent years. Examples

of such devices include Google's Project Glass, Microsoft's HoloLens, and BMW's Mini Augmented Vision. The potential relevance of these sorts of innovations in terms of their ability to support novel forms of cognitive extension has already been described at length in a number of papers (Smart 2012; 2014). In particular, it has been suggested that such devices transform our traditional notions of online information by situating the Web at the heart of our everyday embodied interactions with the world. By virtue of their ability to augment aspects of the real world with digital information, such technologies could easily be seen to support the creation of virtual variants of the designer environments discussed by Clark (2008, 59). Such *virtual designer environments* provide a number of opportunities to reshape the local environment in ways that extend the reach of our own cognitive capabilities. Indeed, by extending the real world with highly malleable digital representations, the profile of our own cognitive endeavors is arguably poised to go beyond the constraints imposed by physical (as opposed to virtual) reality.

In thinking about the contributions of augmented reality technologies to our cognitive capabilities, it is important that we do not think of the notion of virtual designer environments as simply a form of perceptual enrichment of the real world. Although it is clearly true that augmented reality technologies are poised to provide this sort of enrichment, they also provide opportunities for new forms of agent-world interaction. Of particular interest, in this respect, is Microsoft's HoloLens, which aims to enable users to physically interact with virtual representations in the form of holographic projections. The virtual representations, in this case, are ones that are responsive to a variety of physical user actions, and this opens the door to forms of interaction that closely resemble those encountered in philosophical discussions of extended cognition. Imagine, for example, that rather than relying on pen and paper to accomplish the task of long multiplication, a human agent exploited the hologrammatic equivalents of these resources. Inasmuch as the task involving the manipulation of the physical resources (i.e., pen and paper) constitutes a form of extended cognition (see Wilson and Clark 2009), is there any reason to view a task involving the manipulation of hologrammatic representations in a different light? The result, perhaps, is that the advent of hologrammatic computing heralds the arrival of new forms of cognitive extension, such as hologrammatically extended cognition and hologrammatically extended minds.

Thus far, I have said nothing about the extent to which emerging technologies might speak to the criterion of continuous reciprocal causation. This criterion, recall, focuses on the nature of the interactions between the human agent and some bio-external resource. The general idea is that extended cognitive systems are ones in which we encounter reciprocal forms of information flow and influence between the constituent elements of an extended cognitive mechanism (Palermos 2014b).

Now, it might seem that when we focus our attention on the original Otto case we encounter very little in the way of evidence for continuous reciprocal causation. Regardless of whether or not this is true, it is not, I suggest, a criticism that can be leveled at the nature of our interaction with more technologically advanced resources.

The reason for this is that practically every form of interaction we have with a digital device will involve some form of state change, represented by a change in the device's computational activities and/or a change in the informational properties of the device's user interface. Such changes will, in most cases, alter the responses of the human agent, thereby closing the feedback loop. This sort of interactive engagement between device and user is particularly apparent in the case of hologrammatically extended cognition. Here, the actions of a human user lead to an ongoing alteration in the representational and computational activity of the projection device. It might be said that such forms of interactivity are much less apparent in the case of augmented reality displays that merely present information in the user's field of view (e.g., Google Glass). However, even in these cases, there is often a very potent form of reciprocal influence that is taking place. Consider, for example, the relatively simple case of a human user relying on a GPS-enabled augmented reality device to support their navigation to a remote location. The device, in this case, provides a set of graphic affordances that serve to guide the user's movements through the physical environment. As a result of these movements the user is exerting a form of indirect influence on the device, one that causes the device to update its graphical outputs. Despite the indirect nature of the interactive coupling, there does seem to be a cognitively potent form of reciprocal interaction occurring here between the human agent and the technological artifact.

1.4.5 CONFIGURABLE TECHNOLOGIES

The set of criteria outlined by Clark and Chalmers (commonly referred to as the trust+glue criteria) have recently been supplemented with additional criteria, such as the personalization criterion (Clowes 2015; Sterelny 2010). In view of the putative relevance of personalization, it is important to briefly review some of the ways in which digital technologies can be tailored for personal use. Perhaps the most obvious form of personalization that comes to mind here concerns the configuration of device settings and the installation of specific apps and services. In practically all cases, these changes to the default configuration of a device are likely to reflect the specific informational needs and interests of the user, as well as the kinds of task in which the user is likely to engage.

Another form of personalization occurs with respect to the physical configuration of the device itself. This form of personalization is encountered much less often; however, some development initiatives are seeking to develop highly modular devices that can be physically configured (and reconfigured) to optimize the fit between device functionality, user requirements and task context. Perhaps the most high profile of these development efforts is Google's Project Ara.⁴ This initiative aims to create a modular hardware ecosystem that will enable individual users to configure smartphones in ways that are precisely aligned with their functional and aesthetic preferences.

⁴ <http://www.projectara.com/> (last accessed November 24, 2015).

A third form of personalization centers on the use of user profiling techniques, which are sometimes used to monitor the behavior of, e.g., Web users in order to make predictions about their tastes and preferences. Perhaps the best-known example of this sort of personalization comes in the form of the personalized search results that are offered by many search engine providers, such as Google. The aim, here, is to optimize the information retrieval process, putting users in direct contact with information that is likely to be of greatest interest and relevance to them. Such objectives seem to be suitably aligned with at least some of the criteria for cognitive extension, most notably the personalization and accessibility criteria.

1.5 *Web-extended minds, revisited*

1.5.1 TECHNOLOGICALLY EXTENDED OTTO

The foregoing analysis highlights some of the ways in which current research and development efforts are likely to lead to technologies with properties that speak to the criteria for cognitive extension. I have suggested that, in many cases, these properties satisfy the criteria to at least the same extent as the usual philosophical cases. With a better understanding of technological developments now in place, we are in a position to revisit the notion of the Web-extended mind and ascertain to what extent such forms of cognitive extension are enabled by new technologies. In particular, now that we have a better grip on the emerging technological landscape, it is time to give Otto something of an upgrade. Let us therefore meet Otto++.

Otto++

Otto++ is a neurologically impaired individual who is biologically identical to Otto. Otto++ has just purchased a shiny new smartphone and augmented reality glasses. Otto++ spends some time configuring his phone by installing a variety of apps. He then carries his phone with him wherever he goes. In order to ensure that he has access to relevant information, Otto++ installs an app that enables him to record important pieces of information. The app, however, does not store information locally on the device. Instead, it relies on a semantically enabled, cloud-based personal data store that stores information in a linked data format. In order to access his personal data store, Otto++ installs an app that enables him to quickly retrieve important items of information using an intuitive graphical user interface. He also links his phone to his augmented reality glasses so that relevant information from his data store can be presented within his visual field. One day, while on a trip to New York City, Otto++ decides he would like to visit MoMA. He automatically says the word “MOMA” out loud. His phone executes a semantic query against his personal information repository and retrieves information about MoMA. A set of directional indicators appear within Otto++’s visual field, alongside some descriptive information about MoMA.

The case of Otto++ is sufficiently similar to the original case, I suggest, to warrant similar conclusions regarding the extended status of the two protagonists: if Otto

and his notebook constitutes a genuine case of cognitive extension, then I can see no reason why we should not think of Otto++ in similar terms. The nature of the bio-external resources exploited by Otto and Otto++ are, of course, different, but the functional roles of these resources (i.e., their role in guiding thought and action) look to be identical.

The case of Otto++ is important, I suggest, because it enables us to respond to a variety of challenges that have been leveled against the original Otto case. One of these challenges relates to issues of informational updating and inferential integration (Weiskopf 2008; Wikforss 2014). The claim, here, is that Otto's notebook cannot serve as a realization base for his dispositional beliefs because it is a normative constraint on belief states that they are subject to forms of automatic updating that work to ensure consistency with other, semantically related, beliefs. For example, imagine that MoMA is temporarily re-located to Long Island City, as was indeed the case in 2002. The receipt of this information should, according to Weiskopf (2008), result in the automatic revision of whatever beliefs are related to MoMA, e.g., the belief that MoMA is located on 53rd St. The failure to exhibit such forms of updating in the conventional notebook case leads to the suggestion that we should discount the idea that Otto's notebook forms any part of his doxastic machinery.

There are a number of ways in which we might respond to concerns regarding the automatic updating and propagation of belief states (see Clark 2005). Note, however, that in regard to the Otto++ case, it is not entirely correct to say that *no* form of informational updating and inferential integration is taking place. Otto++ relies on a personal data store that avails itself of semantically enriched representations, and, by virtue of this enrichment, the contents of the data store are subject to forms of automatic inference that will include checks on the (logical) coherence and consistency of the stored information. The assertion of new information (such as a change in the location of MoMA) will thus result in a semantically consistent cascade of informational changes, akin to the sort that might be expected to occur in the case of a neurologically normal individual. I, personally, have no problem in casting this as a form of automatic updating of an agent's externally situated belief base. The Otto++ case might thus go some way toward allaying the concerns of Weiskopf (2008) and Wikforss (2014). It should also be clear that in situations where a data repository is located on the Web, it is in an ideal position to monitor information feeds originating from a range of other sources, e.g., institutional websites, social media streams, sensor networks, and so on. In this case, we may assume that Otto++'s data store is subject to continual updates, reflecting the changing nature of reality. Importantly, such updates (providing the various sources of information are reliable) can be expected to yield informational encodings that track states of affairs in the world; i.e., they will reliably track the truth. One advantage of the Otto++ case is thus that it enables us to see how Web-based forms of cognitive extension can enable an agent to track the truth in ways that are unavailable to Otto and his computationally inert notebook. By virtue of exploiting an online

resource that is embedded in a globally distributed informational ecology, Otto++ is in a position to exhibit epistemic performances that surpass those of his technologically low-grade counterpart, Otto, as well as (perhaps) Otto's neurologically normal companion, Inga.

1.5.2 TRUSTING THE WEB: ENDORSEMENT AND SELECTION

At this point, the astute reader may be inclined to highlight a number of problems with the Otto++ case. One problem is that the Otto++ case is not representative of our actual engagements with the Web. Of particular concern is the role that Otto++ plays in the creation of online content. This is not a typical feature of our interaction with the Web. In most cases, we do not create and manage the information that we then go on to exploit; in general, the information is provided by other users, or in some cases, it is automatically generated by remote computers. This turns out to be a crucial difference when it comes to the epistemological analysis of Web-extended cognizers. To see this, recall that the trust criterion mandates the automatic endorsement of bio-externally situated information in extended mind cases. Otto++, we may assume, satisfies this criterion; i.e., he automatically endorses whatever information he retrieves from the online environment. From an epistemic perspective, the consequences of automatic endorsement look to be relatively benign in this case: Barring some form of outside interference—a security breach, for example—it seems reasonable to assume that many of Otto++'s beliefs will reliably track the truth—Otto++ is, after all, the one who curated much of the content for his personal data store. Things are very different, however, when we consider a more realistic scenario involving the exploitation of Web-based content. In this case, we are interacting with a highly public and shared information space, one in which much of the content is created by other agents and misinformation and mendacity are ever-present hazards. In this situation, the epistemic consequences of automatic endorsement are, to say the least, worrying. An agent who automatically endorses everything they encounter online is likely to harbor more than a few false beliefs. In fact, we might go so far as to say that any true beliefs that the agent does hold in this situation are only true as a result of some fortunate happenstance—the fact that the agent stumbled across the right resource at the right time. Given that luck is almost universally seen as inimical to knowledge in mainstream analytic epistemology (Pritchard 2009; 2012; Riggs 2007), our chances of encountering real-world agents that function as both Web-extended cognizers and Web-extended knowers (i.e., agents that simultaneously fulfill the criteria for cognitive extension and knowledge attribution) may look a little bleak.

The trust criterion thus presents something of a problem when it comes to our normal pattern of engagement and interaction with the Web. How could an agent confronted with the informational equivalent of the Wild West possibly hope to know anything in the absence of a fair amount of epistemic vetting, checking, and selective endorsement? And if automatic endorsement is off the table, then how could any Web-extended

cognizer also count as a Web-extended knower? To make matters worse, it is not entirely clear that Web-extended cognizers will be all that plentiful. As Clark (2010b) points out, the nature of the online environment seems to count against the easy satisfaction of the trust criterion. In particular, Clark (2010b) suggests that certain forms of Web access (e.g., mobile access to Google) are unlikely to yield information that is trusted to the same extent as information retrieved from biological memory. As a result, it seems that actual cases of Web-extended cognition may be something of a rarity.

Given that we have probably gone as far as we can with the Otto++ case, let's introduce another case to help inform our consideration of trust-related issues.

Jesse

Jesse is a neurologically normal individual who has just purchased a shiny new smartphone and augmented reality glasses. Like Otto++, he engages in the initial configuration of his phone by installing a variety of apps, and like Otto++ he carries his phone with him wherever he goes. Unlike Otto++, however, he does not attempt to manage a personal repository of semantically enriched data; instead, he relies on a number of linked data services that already exist on the Web. In order to make use of these services he installs an app that enables him to quickly retrieve important items of information using an intuitive graphical user interface. He also links his phone to his augmented reality glasses so that relevant information can be presented within his visual field. One day, while on a trip to New York City, Jesse decides that he would like to visit MoMA. He automatically says the word "MOMA" out loud. His phone executes a semantic query against a number of linked data services in order to retrieve information about MoMA. A set of directional indicators subsequently appear within Jesse's visual field, alongside some descriptive information about MoMA.

From an epistemic perspective, at least, it should be clear that Jesse is not in a position where he can afford to automatically endorse the deliverances of his Web-enabled device—such information, we may assume, is subject to a fair amount of epistemic pollution and contamination. Ideally, Jesse should be able to evaluate the information he receives before he endorses it. The problem, of course, is that this sort of evaluative effort seems to fly in the face of the trust criterion.

It is important to note, here, that there are actually two possible readings of the trust criterion, especially when it comes to the notion of automatic endorsement. The first reading (AE_{STRONG}) is that externally situated information should not be subject to *any* form of evaluative assessment. The second reading (AE_{WEAK}) is that the information should not be subject to a form of evaluation that is unlike that seen in the case of information retrieved from bio-memory. I suspect that what Clark (2010b) is trying to guard against by invoking the trust criterion is a form of evaluative assessment that is unlike that encountered in the case of internally situated information flows. In view of this, I suspect that Clark would be happy to accept the second reading of the trust criterion (i.e., AE_{WEAK}). It should be clear that, at least from an epistemic perspective, AE_{WEAK} is also the preferred reading. If we were obliged to commit to AE_{STRONG} , then it

would seem that in order for Jesse to be a Web-extended agent he would always be required to believe the information he retrieved from the Web. Such a commitment would mean that Jesse could exercise no flexibility with regard to the endorsement of externally derived information (i.e., he is committed to a form of doxastic involuntarism). If this is indeed the case, then it seems unlikely that Jesse could exist as a Web-extended agent of the epistemically virtuous variety.

Things look much brighter, of course, if we reject AE_{STRONG} in favor of AE_{WEAK} . In this case, it seems that we can allow for the possibility of some form of evaluative assessment. In fact, one reason to accept AE_{WEAK} derives from what is actually intended as a critique of the extended mind thesis by Michaelian (2012a). Michaelian's argument is grounded in the reconstructive nature of bio-memory (Hassabis and Maguire 2007; Matthen 2010; Schacter and Addis 2007). He suggests that by virtue of reconstruction we do not automatically endorse information from bio-memory. Instead, we rely on a sub-personal endorsement mechanism that works to filter information as part of the process of belief formation. The operation of this endorsement mechanism in the case of bio-memory, Michaelian (2012a) argues, is sufficient to exclude cases of technology-mediated information retrieval as counting as a form of extended memory on the grounds that the extended mind theorist is committed to the notion of automatic endorsement.

It should be relatively clear that something has gone awry here, since Michaelian seems to be arguing against the possibility of cognitive extension in a way that services our needs with respect to the possibility of extended knowledge. The root of the confusion, I suggest, lies in the different readings of the trust criterion. Michaelian's (2012a) critique is primarily directed toward the stronger reading of the trust criterion, i.e., AE_{STRONG} . If the extended mind theorist is indeed committed to AE_{STRONG} , and if Michaelian is right about endorsement in the case of bio-memory, then there is clearly an issue that is deserving of further thought and reflection. If, however, the extended mind theorist is only committed to the weaker reading of the trust criterion then much of the sting seems to be taken out of Michaelian's (2012a) critique.⁵

My own view here is that there is something compelling about Michaelian's claims regarding endorsement, and this, I suggest, should encourage extended mind theorists to embrace the weaker reading of the trust criterion (i.e., AE_{WEAK}). Not only does this move seem to be warranted by the reconstructive nature of biological memory (as Michaelian suggests), it also helps to highlight the functional role of endorsement mechanisms in enhancing the reliability of a belief-forming system. In particular, if we treat endorsement as a form of metacognitive process that selectively gates the control that information (inner or outer) has over an individual's thoughts and actions (which is, presumably, the basis for belief ascription), then we can model belief formation as

⁵ If Michaelian (2012a) is wrong about biologically based endorsement, of course, then it does not matter what reading of the trust criterion the extended mind theorist chooses to embrace: in either case, the extended mind thesis comes through the attack unscathed.

a two-level process involving both a first-order process, which is responsible for the acquisition of belief-relevant information, and a second-order process, which monitors the activities of the first-order process (Michaelian 2012b). This makes the endorsement mechanism a form of metacognitive (second-order) process that plays a productive role in determining the overall reliability of the belief-forming system in which the metacognitive process is situated. As a means of getting a feel for the nature of this influence, consider a situation in which some first-order belief-forming process (the one that gathers or acquires information) is perfectly reliable. In this case, the second-order process looks to be redundant—it cannot enhance the reliability of a process that is already 100 percent reliable. This does not mean, however, that the second-order process is irrelevant when it comes to the reliability of the larger belief-forming system. The reason for this is that even if the first-order process were to be 100 percent accurate, the second-order process would still need to deliver accurate estimates of the reliability of the first-order process (i.e., the second-order process would itself need to be reliable). Imagine a situation where the first-order process is 100 percent accurate, but the second-order process accurately estimates the reliability of the first-order process only 50 percent of the time. In this case, the reliability of the total system (comprising both first- and second-order processes) will be reduced by half. For this reason, it is important that the second-order process reliably estimates the reliability of the first-order process.⁶

In addition to issues concerning endorsement, Michaelian (2012b; 2014) also talks about what he refers to as the “selection problem.” This is characterized as “the problem of selecting one’s resources so that they provide one with answers to whatever questions are currently driving inquiry, taking into account the availability of resources, their reliability, the costs involved in their use, and so on” (Michaelian 2014, 321). The selection problem is thus the problem of selecting an appropriate cognitive strategy in a given context. In the case of extended cognition, the primary problem confronting the agent relates to the choice of strategies involving bio-internal (e.g., bio-memory) or bio-external (e.g., a notebook) resources. This problem seems to be of particular relevance in the case of Jesse, since, in addition to selecting the resource (i.e., the portable device), Jesse may also need to choose between a range of apps and services that provide access to information of different kinds and different quality. A tendency to engage in the selection of particular sources of information, especially those that provide accurate and reliable information, is likely to play a key role in Jesse’s propensity to form true beliefs about the world. Selection, as with endorsement, can thus be seen as a form of epistemically virtuous metacognitive activity, one that reflects an agent’s cognitive abilities and agency. This is explicitly recognized by Michaelian (2014) when he suggests that “cognitive agency can contribute to the selection of

⁶ When it comes to the problem of selecting between competing resources or cognitive strategies, we might argue that the second-order process needs to reliably *predict* the reliability of the first-order process. This seems to be a second way in which the reliability of a belief-forming system could be enhanced.

the relevant resources, the assembly of the distributed [i.e., extended] system, and the endorsement of information produced by the system” (320–1). (The significance of such claims will become apparent later on when we examine the role played by cognitive agency in virtue reliabilistic approaches to knowledge attribution.) Also note that by linking the selection process to the relative reliability (i.e., accuracy) of different information sources, we might go some way toward addressing concerns about the original (trust-related) endorsement problem. In fact, inasmuch as Jesse is selecting processes based on their tendency to yield reliable information, then we might wonder whether Jesse is still required to engage in the kinds of epistemically hygienic practice associated with AE_{STRONG} ; i.e., the tendency to subject information to conscious scrutiny and deliberate evaluation. Perhaps all that is required in this situation is for Jesse to be responsive to epistemically relevant factors that indicate errors regarding the predicted reliability of an extended belief-forming process. In fact, by being suitably responsive to counterfactual error possibilities, it seems that Jesse can simply endorse the informational outcomes of whatever process was selected (providing that he has no reason to doubt the accuracy of his initial reliability-related estimates). Here, we can see that much of the emphasis concerning epistemically relevant capabilities is being shifted to the selection phase of a belief-forming process. Inasmuch as Jesse selects processes according to their predicted reliability and (let us assume) his prediction estimates are always accurate, then endorsement appears to be of little relevance with regard to his subsequent epistemic standing.

In summary, in order to allay concerns about the trust criterion and the extent to which it excludes the possibility of Web-extended minds, I suggest it helps to reflect on the following points:

1. Epistemic agents have the opportunity to configure and personalize their devices in ways that make them (the devices) more reliable and trustworthy. This may minimize the extent to which the agent needs to subject the informational deliverances of a device to the kinds of critical scrutiny that Clark finds so problematic.
2. The notion of automatic endorsement does not rule out the possibility of some form of sub-personal information evaluation taking place. All that is required is that we don't see some radical departure from the kind of vetting and checking that goes on in the case of biological memory (or, more generally, internally situated information flows).
3. Epistemic agents may select belief-forming processes on the basis of their predicted reliability. If a process is selected because it is reliable, then active scrutiny and conscious endorsement may be of nugatory significance with regard to issues of epistemic standing. Importantly, the epistemically virtuous selection of belief-forming processes (extended or otherwise) seems to allow us to accept the trust criterion (or some variant thereof) without thereby denying the possibility of Web-extended knowers.

1.5.3 METACOGNITION AND THE NETWORK-EXTENDED MIND

Michaelian (2012b; 2014) suggests that we should see endorsement and selection as examples of metacognitive processes, i.e., as processes that are concerned with the monitoring and control of other cognitive processes. The role played by these mechanisms in strategy selection and informational endorsement establishes contact with an important idea in the philosophical and cognitive science literature regarding the adaptive construction and configuration of cognitively relevant information-processing circuits. This idea is apparent in the so-called *principle of ecological assembly*, according to which “the canny cognizer tends to recruit, on the spot, whatever mix of problem-solving resources will yield an acceptable result with a minimum of effort” (Clark 2008, 13). It is also apparent in work that seeks to assess the relevance of network scientific and information theoretic approaches to the analysis of extended cognitive systems (Smart et al. 2010). Smart et al. (2010) thus suggest that selection processes contribute to the creation of dynamically configured (or dynamically assembled) networks that alter the information-processing profile of some larger cognitive organization. It may help here to think of the Otto case in network-oriented terms. Thus, when we think of Otto reaching for his notebook, perhaps we should view his hands as something like the electrical contacts of a switch that, when they come into contact with the notebook, serve to establish a new kind of (extended) electrical circuit. The potential of biological neural networks to be subject to this kind of dynamic configuration has been reported by studies investigating the processing dynamics of invertebrate nervous systems (Meyrand et al. 1994). It is also implied by the notion of temporary, dynamically assembled task-specific devices as discussed by Anderson et al. (2012). There is, moreover, considerable interest in understanding how the so-called *effective connectivity* of a socio-technical ensemble can be dynamically modified in order to meliorate cognitive processing in distributed or collective cognitive systems (e.g., Smart et al. 2014). One study that is of particular interest here is a study by Reitter and Lebiere (2012), which helps to shed light on the role of metacognition in dynamically configuring the structure of a communication network in a team-based problem-solving task. On the basis of a number of computer simulation experiments, Reitter and Lebiere (2012) conclude that the “metacognitive adaptation of communicative behavior can optimize the spread of information through a communication network” (243). Such conclusions suggest that metacognitive processes may play an important role in determining the effective connectivity of information-processing networks at both the level of individual cognitive agents and larger collective cognitive ensembles, such as collective doxastic agents (Palermos and Pritchard 2013; Palermos 2015) and socio-epistemic systems (Goldman 2011).

With all this in mind, we can begin to think about the relevance of changes in the connectivity of an information-processing network relative to the epistemic standing of cognitive agents. Hints that such changes might be of epistemological significance are apparent in a recent paper by Clark (2015). Clark (2015) suggests that neural processes of precision estimation and precision weighting might work to influence our responsivity

to the context-varying reliability of inner and outer sources of information. In addition, he suggests that the same processes might support moment-to-moment changes in the structural organization of network-mediated information flows. Inasmuch as we see precision estimation and precision-weighting mechanisms as contributing to the realization of sub-personal metacognitive processes, then Clark's (2015) analysis helps us see metacognitive processes as playing an important role in determining the *effective* structure of information-processing networks that ground our epistemically virtuous engagements with bio-external resources. Such a view seems to be endorsed by at least some members of the epistemological community. When it comes to the issue of determining whether an agent should be credited with knowledge, for example, Proust (2014, 386) suggests that we should ask:

...how normatively appropriate (from a rational-instrumental and an epistemic viewpoint) the selection, control and monitoring of the mediators have been, in a particular epistemic action or set of actions, with respect to the goal of producing true beliefs.⁷

The idea, then, is that when it comes to cases of extended cognition, sub-personal metacognitive processes seem to be ideally placed to determine the overall shape of an agent's information-processing economy. Moreover, these processes seem to be of potential epistemic significance, contributing to the selection of reliable information sources and determining the extent to which those sources can influence the expression of behaviors that warrant the ascription of beliefs. In the next section, I will attempt to apply these ideas to the notion of extended knowledge. I will also attempt to examine the impact of these ideas on claims that a particular kind of epistemological theory is compatible with active externalist approaches to mind and cognition (see Pritchard 2010).

2. Extended Knowledge

2.1 *Virtue reliabilism*

One of the driving forces behind recent work in epistemology is the intuition that an agent's true beliefs should only count as knowledge if the beliefs are linked to the exercise of an agent's cognitive abilities. This intuition, commonly referred to as the ability intuition (Pritchard 2009), lies at the heart of a popular theory of knowledge that goes by the name of virtue reliabilism. A key feature of virtue reliabilistic accounts is the

⁷ This may be the reason why we are reluctant to ascribe knowledge to the unfortunate protagonist in the Alvin case (see Pritchard 2010). Here, we have an agent—Alvin—who has a brain lesion that causes him to form the belief that he has a brain lesion. We are somewhat reluctant to credit Alvin with knowledge, I suggest, because there is no evidence of the sort of reliability-enhancing mechanisms (i.e., selection, monitoring, and endorsement) that we encounter in typical cases of ecologically normal belief formation. In particular, Alvin has no control over which belief-forming process is selected, and he has no opportunity to gate the control that (internally derived) information has over his subsequent thoughts and actions.

emphasis they place on cognitive abilities in determining the truth status of an agent's beliefs. In general, virtue reliabilists see knowledge as a form of cognitive success (i.e., true belief) that is grounded in the integrated set of cognitive abilities that constitute the agent's cognitive character.⁸ Knowledge is thus seen as a form of cognitive success that stems from the exercise of cognitive ability.

There are a number of ways in which the basic virtue reliabilistic position might be fleshed out. The two forms of virtue reliabilism that will occupy us here are those that have attracted the most attention in terms of the issue of cognitive extension (Kelp 2013; Palermos 2011; 2014a; 2015; Pritchard 2010). The first of these positions is known as robust virtue reliabilism. Pritchard (2010) characterizes this form of virtue reliabilism as follows:

Robust Virtue Reliabilism

S knows that *p* iff *S*'s true belief that *p* is the product of a reliable belief-forming process which is appropriately integrated within *S*'s cognitive character such that her cognitive success is primarily creditable to her cognitive agency.

One of the important things to note here is the emphasis on primary creditability. In particular, robust virtue reliabilism entails a commitment to the idea that knowledge should be linked to an agent's cognitive character and cognitive agency such that cognitive success (i.e., believing the truth) is something for which the agent deserves *primary* credit. In other words, it is an agent's cognitive abilities that figure most prominently in causal explanatory accounts as to why it is that an agent believes the truth.

An alternative to robust virtue reliabilism comes in the form of the "weak cognitive agency" (COGA_{WEAK}) view proposed by Pritchard (2010):

COGA_{WEAK}

If *S* knows that *p*, then *S*'s true belief that *p* is the product of a reliable belief-forming process which is appropriately integrated within *S*'s cognitive character such that her cognitive success is to a significant degree creditable to her cognitive agency.

COGA_{WEAK} differs from robust virtue reliabilism in two ways. First, the commitment to the biconditional in the case of COGA_{WEAK} has been dropped. This move is deemed to be necessary in order to accommodate cases of "environmental luck" in which an agent's beliefs end up being true by virtue of some quirk of the belief-forming environment (see Pritchard 2012). The second difference between robust virtue reliabilism and COGA_{WEAK} relates to the degree to which cognitive success is creditable to cognitive agency. In the case of robust virtue reliabilism, we encountered the notion of primary creditability. COGA_{WEAK} reduces the extent to which the truth of a belief is creditable to agential abilities by invoking the notion of *significant* credibility. It is this weakening of the credibility constraint that is deemed to make COGA_{WEAK} particularly well suited

⁸ An agent's cognitive character is defined as "her integrated web of stable and reliable belief-forming processes" (Pritchard 2010, 136).

to accommodate cases of extended cognition and extended knowledge. This is because cognitive extension is often seen to entail a distribution of credit relative to what it is that an agent believes, with some of the credit being attributable to non-biological elements.⁹ Inasmuch as we accept this claim, then it seems that an emphasis on primary creditability is likely to cause problems in terms of the attempt to attribute knowledge to extended cognizers. In particular, it seems that, in some cases, what an agent believes may have more to do with the properties of some bio-external resource than it does with the properties of the biologically bounded individual. The claim, in essence, is that a weakened form of virtue reliabilism, in the form of COGA_{WEAK}, establishes something of a “snug fit” with respect to active externalist accounts (Pritchard 2010). Such snugness is surely a source of comfort when it comes to understanding how virtue reliabilism and active externalism might work together to inform our understanding of extended knowledge.

2.2 *Web-extended knowers*

Having discussed the ways in which emerging digital technologies might support extended cognition, and with some sense as to how contemporary epistemologists view the notion of knowledge, it is time to consider the ways in which technologically advanced forms of cognitive extension might contribute to positive epistemic standing (at least from the standpoint of virtue reliabilistic accounts of knowledge). In this section, I will attempt to examine the notion of extended knowledge and relate it to the earlier discussion on Web-extended minds. I will also attempt to assess the relationship between extended cognitive processes, cognitive ability, and epistemically relevant credit attributions.

2.2.1 EXTENDED KNOWLEDGE

Given the idea that knowledge is characterized as a form of belief in mainstream epistemology, one way of interpreting the term “extended knowledge” is to see it as simply referring to a state of extended belief, i.e., as a belief that is materially extended by virtue of the role that non-biological (or extra-organismic) elements play with respect to its physical realization. Clearly, this view sits very comfortably alongside claims regarding the possibility of extended realization bases for states of (dispositional) belief; i.e., the claims of the extended mind theorist. Given the possibility of extended minds, it should, in principle at least, be possible for agents to enjoy states of extended knowledge. All that we need to add to the basic extended mind account, it seems, are some additional conditions that work to ensure that at least some of an agent’s extended beliefs qualify as knowledge. And, with our virtue reliabilistic account of knowledge now to hand, it seems relatively clear what these conditions are. In particular, what needs

⁹ Clark and Chalmers (1998), for example, claim that cases of cognitive extension entail the distribution of epistemic credit. “Epistemic action,” they suggest, “demands spread of epistemic credit” (8). See Preston (2010) for a critical analysis of this claim.

to be added in order for an extended believer (*S*) to count as an extended knower is that *S*'s extended beliefs should be 1) true and 2) the product of belief-forming processes that reflect the exercise of *S*'s cognitive abilities.

When we apply all this to the cases of Otto++ and Jesse, however, we have a problem. Note that when we talk about Otto++ and Jesse's extended minds what we are really talking about are states of dispositional belief, as opposed to states of occurrent belief. The belief-forming processes of interest are thus those that contribute to the *formation* of states of dispositional believing. In the case of Otto++, this is not necessarily a problem—Otto++, recall, deliberately manufactures and maintains his extended belief base, and we may assume that he exercises considerable ability in ensuring that such beliefs reliably track the truth. This is not the case with Jesse. Although Jesse may exhibit a lot of epistemic virtue in configuring his Web-enabled device, ensuring that it functions properly, and that it delivers reliable information, he seems to play little or no active role in *forming* the beliefs that are housed within his bio-external store. In fact, it seems highly likely that it is some other entity (human or machine) that is responsible for the formation of the (dispositional) beliefs that ultimately get ascribed to Jesse. Assuming that we want to treat extended dispositional beliefs as the basis for states of extended knowing, then it looks as though virtue reliabilism gives us different answers regarding the possibility of extended knowledge in the Otto++ and Jesse cases: attributions of extended knowledge seem appropriate in the case of Otto++, but they seem entirely inappropriate in the case of Jesse.

Perhaps one way of responding to all this is to insist that it is not the content of the focal belief that matters, so much as its truth status. Why is it that Jesse's dispositional beliefs are true? Answer: because of the way in which Jesse has configured his device so as to ensure the delivery of reliable information. Here, Jesse's abilities in configuring his device seem to play a key role in terms of our attempts to understand why it is that Jesse believes the truth, and this obviously serves as a potential point of contact with virtue reliabilistic theories. Note, however, that this response places a lot of emphasis on actions that need not always take place. What if Jesse is the recipient of a new, pre-configured device that is so well designed as to not require any form of epistemically motivated tinkering—the factory settings, we may assume, are perfectly adequate for Jesse's epistemic purposes. In this situation, it seems that Jesse is deserving of very little credit regarding either the content of his dispositional beliefs or their truth status.

An alternative response might be to insist that what really matters in this situation is the way in which Otto++ and Jesse retrieve information from their respective devices. What really matters, we might say, are the processes associated with information source selection, device interaction and (possibly) the selective endorsement of belief-relevant information. I am inclined to feel that there is some merit to this view, not least because the notion of cognitive extension seems best suited to those situations in which an external resource is actively contributing to the realization of cognitive processes in the here-and-now, as opposed to situations in which some external resource is being created and/or configured to serve some future cognitive purpose. The problem, however, is

that by discounting all the processes that contribute to the *formation* of an agent's dispositional beliefs and focusing our attention solely on the run-time processes, we open ourselves up to the criticism that the relevant run-time processes (even though they may be extended and even though they may very well reflect the exercise of an agent's cognitive abilities) are not so much contributing to the formation of an extended dispositional belief as they are to the formation of a non-extended occurrent belief.

We thus encounter a dilemma: either we focus on the processes that contribute to the formation of a body of extended dispositional beliefs (which, relative to virtue reliabilistic theories, seems fine for Otto++ but not for Jesse), or we focus exclusively on the processes that operate in the here-and-now (the ones that shape the course of an agent's actual behavior) and face the charge that the resulting belief states are not extended; i.e., all the vehicles that contribute to the material realization of the focal beliefs are contained within the head of the target agent. In either case, it seems that we have a problem in applying the notion of extended knowledge to Jesse. On the one hand, attributions of extended knowledge seem to be unwarranted on account of the fact that Jesse has not contributed to the formation of his dispositional beliefs. This encourages us to focus our attention on the actual processes that are in play when Jesse retrieves information from an external resource.¹⁰ The problem we then have, however, is that we make ourselves vulnerable to the charge that all Jesse really has (in terms of his epistemic standing) is a set of non-extended, occurrent beliefs. In either case, we do not seem to confront a situation where any form of extended knowledge (in the form of extended belief states) is in evidence.¹¹

Perhaps the way out of this dilemma is to link notions of dispositional believing to the nature of the functional poise that exists between a biological agent and their technological accoutrements. Something along these lines is, in fact, suggested by Clark (2010a) when he claims that “the very notion of a dispositional belief already makes implicit reference to what would happen in possible run-time situations” (88). The idea, then, is that the functional poise of the externally situated information—its ability to guide thought and action in the manner of brain-bound, non-extended belief states—is sufficient to warrant talk of the agent as already believing such and such, even before the cognitive circuit, incorporating the external resource, is actually established.

¹⁰ Just to be clear, the claim is not that Jesse is exercising *no* ability in this situation, and thus that his beliefs are not, in some sense at least, creditable to him. We could still claim that Jesse needs to endorse the deliverances of his device in an epistemically virtuous manner. My claim is merely that, if we did highlight the role played by endorsement, it seems that we are emphasizing the role of a brain-bound process in contributing to the truth status of an occurrent (i.e., non-extended) belief. Also note, by way of an aside, that if we eliminate the role of endorsement here and assume that Jesse is fed a steady stream of accurate information from his pre-configured device, then the role of Jesse's agency in securing the truth status of his beliefs seems to shrink to such an extent that it appears Jesse should be afforded no credit in securing the truth of his beliefs.

¹¹ Note that this sort of problem only comes about as a result of the emphasis accorded to belief-forming processes in securing positive epistemic status in the context of virtue reliabilistic conceptions of knowledge. This is not necessarily a dilemma faced by extended mind theorists who limit their attention to extended states of dispositional belief.

The problem, however, is that it is not immediately obvious (in the case of Jesse, at least) that we have any sort of process in play (other than those associated with the establishment of the functional poise) that actually contributes to the formation of the *dispositional* state of believing (or knowing). The only processes that do seem to be in play when it comes to the formation of Jesse's dispositional beliefs are those that involve other agents (the ones who create the online content that Jesse subsequently exploits). Again, if we switch our attention to the run-time processes that exploit the poise of bio-external information then it is unclear to what extent we are still talking about processes involved in the realization of *extended* belief states. It seems much more appropriate, in this case, to think that we are talking about processes that contribute to the formation of *non-extended* occurrent beliefs.

2.2.2 EXTENDED BELIEF-FORMING PROCESSES

Given the difficulty in tying the notion of extended knowledge to extended realization bases for states of dispositional belief, perhaps we should attempt an alternative reading of the term “extended knowledge.” The alternative reading comes in the form of the claim that all that matters in the case of extended knowledge is that an agent possesses knowledge in virtue of having formed a belief on the basis of an extended cognitive process, one which reflects the exercise of an agent's cognitive abilities. This claim seems perfectly acceptable, since nothing in the virtue reliabilistic conception of knowledge commits us to the idea that states of knowledge need to be materially extended. As a result, the change in focus—from extended states to extended processes—seems perfectly compatible with virtue reliabilism. In addition, this change of focus is consistent with the approach taken by many epistemologists who have sought to investigate the relation between virtue reliabilism and active externalism (Palermos 2011; 2014a).

So, armed with this alternative view of extended knowledge, we can return to the issue of whether Jesse and Otto++, by virtue of their technologically mediated interactions with the Web, are able to serve as role models for what we might call Web-extended knowers. At first sight, a positive response to this question looks to be in the offing. Both Otto++ and Jesse can thus be said to participate in extended cognitive processes (even if we restrict the processes of interest to those associated with the retrieval of Web-accessible information), and, providing such processes are reliable, it seems that both Otto++ and Jesse are poised to enjoy the potential epistemic explosion that is envisaged by the likes of Ludwig (2015) and others.

On closer examination, however, another problem emerges. The problem, in this case, is not that Otto++ and Jesse are unable to enjoy a potential explosion of knowledge, nor is it the case that the deployment of extended cognitive processes is irrelevant to explanations as to why this explosion might occur. Rather, the problem is that it is far from clear that it is the extended process itself that is doing all of the explanatory work in accounting for why Otto++ and Jesse believe the *truth*. Virtue reliabilists are relatively clear on this point. As Palermos (2011) comments, “the claim is that the

cognitive success [i.e., true belief] must be a product of a reliable cognitive belief-forming process. It is not the weaker claim that the cognitive ability must have been involved in the acquisition of one's true belief, since this can be satisfied far too easily in ways that do not exclude luck" (745). Similarly, Greco (2010) writes that "[i]n cases of knowledge, *S* believes the truth because *S* believes from intellectual ability – *S*'s believing the truth is explained by *S*'s believing from ability" (75).

The question, therefore, is to what extent do extended cognitive processes explain the truth status of Otto++ and Jesse's beliefs? Is it the extended cognitive processes that explain why their beliefs are true... or is it something else?

In order to answer this question, let us first focus on the case of Otto++. Otto++, we may assume, only ever stores factually accurate information in his personal data repository—information that he has assessed as relevant to his epistemic endeavors and which he believes to be correct. Here, the issue as to whether Otto++ deserves credit for the truth status of his beliefs seems relatively clear: Otto++ is the one who carefully curated the information that went into his online data store, so the reason why Otto++ believes the truth has something to do with his curatorial capabilities, ones that we may assume involve a fair amount of epistemic discernment and cognitive ability. But what about Jesse? Jesse, recall, has to deal with something akin to the epistemic Wild West; he has to cope with the fact that he may be misinformed, manipulated, and deceived. He deals with this problem by bringing his various metacognitive capabilities to bear on the problem, adaptively selecting the most reliable information resources on offer and responding to a variety of epistemically relevant factors via an endorsement mechanism that selectively gates the control that externally derived information has over his thoughts and actions. But what is it that accounts for the fact that Jesse believes the truth? The answer to this question, to my mind, has to be the various metacognitive processes that are in play when it comes to the selection of the (technologically extended) cognitive strategy, the monitoring of its execution, and the evaluation of its informational products. For what else could explain why Jesse believes the truth? It might be the fact that the information resources Jesse relies on are highly accurate and reliable. However, Jesse was not obliged to exploit these sources of information. His selection of those resources comes about as a result of his abilities to select and discriminate between competing sources of information on account of their relative accuracy, reliability, and relevance to his epistemic needs and concerns. Similarly, Jesse should not believe everything he encounters online. If he did, then any true beliefs he formed would say more about the epistemic safety of the online environment or his luck in hitting upon a particular body of information than it would about his cognitive abilities. It seems, therefore, as though the sub-personal metacognitive mechanisms of selection, monitoring, and endorsement are playing an explanatorily crucial role when it comes to the issue of why Jesse believes the truth.¹² This is not to say

¹² Something along these lines is also suggested by Palermos (2015). He comments that "[i]t is the subject's organismic [i.e., biological] cognitive faculties that are first and foremost responsible for the recruitment,

that what Jesse believes has nothing at all to do with the nature of his extended cognitive processes. The extended processes may play an explanatorily significant role in determining what it is that Jesse believes, i.e., the content of his belief. However, whether such beliefs line up with the facts (i.e., whether they track the truth) is something that arguably relates to more than just the nature of the extended process itself. What we seem to require, therefore, is a broader explanatory account, one that acknowledges the role played by the mechanisms that are involved in the selection of a particular process, the monitoring and adjustment of its execution, and the endorsement (or rejection) of its informational outcomes. The abilities that thus seem to be important, in terms of ascribing knowledge to Jesse, are ones that are typically glossed as metacognitive in nature—they are the abilities that seem to form part of what we might refer to as Jesse's *metacognitive character*.

One reason to think that these metacognitive processes are important, specifically in the Jesse case, has to do with the variable reliability of the extended cognitive processes that Jesse deploys to access online information. As with most real-world processes, the reliability of the extended process (i.e., its propensity to deliver accurate information) is likely to vary according to the context in which it is invoked. Jesse's extended belief-forming process may be generally reliable as a process type, based on multiple token instantiations of the process over time. But its actual reliability in specific situations is likely to vary considerably, especially when the process in question draws on resources that are embedded in the technological and informational ecology of the Web. In this case, the "contingent reliability" (for want of a better term) of the process is not same as its "general reliability." As a means of dealing with context-specific variations in contingent reliability, Jesse will need to be sensitive to a range of factors that indicate the reliability of whatever extended process has been deployed. He may pick up on cues that are provided by the information source itself, or he may detect inconsistencies with information retrieved on prior occasions and stored in bio-memory. Whatever the factors that determine the contingent reliability of the extended process, Jesse will need to exhibit some degree of responsivity to these epistemically relevant factors. Most plausibly, the basis for that responsivity lies in the operation of metacognitive processes, which may themselves be based on the predictive processing capabilities of the biological brain (see Clark 2015).

We thus come face-to-face with an important question: Given that metacognitive processes seem to bear much of the explanatory weight when it comes to understanding why it is that an agent, such as Jesse, believes the truth, to what extent are such processes subject to technological extension? The answer to this question is relevant to the issue of whether a weakened form of virtue reliabilism (*viz.*, COGA_{WEAK}) establishes something of a snug fit with regard to the notion of active externalism (Pritchard 2010). If the metacognitive processes are not, typically, extended, then there

sustaining and monitoring of the extended belief-forming process...in virtue of which the truth...is eventually arrived at" (2963).

seems little reason to accept the claim that there is a distribution of epistemic credit in cases where knowledge results from the operation of an extended cognitive process. It is this distribution of epistemic credit, recall, that motivates the rejection of robust virtue reliabilism in favor of the COGA_{WEAK} alternative.

2.2.3 METACOGNITION, EPISTEMIC CREDIT, AND PREDICTIVE PROCESSING

Inasmuch as the foregoing role of metacognitive processes in establishing the truth status of an agent's beliefs is anywhere near the mark, then it seems that the move to COGA_{WEAK} relies on the extent to which metacognitive processes can be cast as processes that supervene on physical vehicles that are located beyond the organismic boundary. It is here that I fear COGA_{WEAK} may be in some trouble. For I suspect that the sort of metacognitive processes that we have been discussing are not ones that are, in most cases, extended. Instead, they seem to be processes that are most likely to be implemented in the neural wetware of the brain, at least in the case of individual epistemic agents. The best way of understanding such processes, to my mind, is along the lines of the predictive processing account described by Clark (2013; 2016). In fact, in a recent paper, Clark (2015) proposes a link between neurally grounded predictive processing mechanisms and metacognitive capabilities. The kinds of processes in question are thus ones that are, for the most part, housed within the biological brain. Indeed, it is not immediately obvious to me how the mechanisms at work in metacognition might be subject to technological or social extension,¹³ at least in the case of individual cognitive agents. This is not to say, however, that such processes cannot be *influenced* by a variety of bio-external goings-on. As Clark (2015) usefully notes "... precision estimation itself can be partially outsourced, scaffolded, and amplified using bio-external tools and resources" (3771).

What, then, of the claim that there is an inherent compatibility between a weakened form of virtue reliabilism (i.e., COGA_{WEAK}) and active externalist accounts. My view on this matter is that, in light of the foregoing discussion, we should not accept the conclusion that there is an inherent conceptual tension between virtue reliabilistic accounts (of any stripe) and active externalism. It seems perfectly possible, to my mind, that Otto++ and Jesse could participate in all manner of extended cognitive processes, and, by virtue of their metacognitive abilities, they would still be deserving of primary credit for their true beliefs. This is not to say, of course, that Jesse and Otto++ do not stand to benefit, in an epistemic sense, from their cognitive incorporation of bio-external resources—we could still witness the much-vaunted explosion of knowledge that Ludwig (2015) anticipates as a result of Web-based forms of bio-technological merger. All that is being claimed here is that we should see Otto++ and Jesse as deserving of *primary* credit (or blame) for what they believe in these situations.

¹³ Kelp (2014), in fact, presents a case—the Timekeeper case—where the activities of another agent play a role that is described as being functionally equivalent to that accomplished by internally situated monitoring mechanisms.

Where does this leave us in terms of our conception of extended knowledge and extended knowers? I suggest that a useful approach is to see the terms “extended knowledge” and “extended knowers” as applicable to situations in which some form of extended cognitive process is relevant to the explanation of why an agent has the beliefs they do, irrespective of whether or not those processes are the ones that ultimately determine the truth status of the beliefs in question.

2.3 Extended cognizers vs. extended knowers: a technology-induced tension?

Finally, what can we say about the impact of emerging digital technologies on our epistemic profile? Are we finally in a position to embrace the claims of Ludwig (2015), Bjerring, and Pederson (2014), and others regarding the expansion of our epistemic capabilities in the wake of Web-based forms of cognitive extension? Unfortunately, for a variety of reasons, I do not think things are quite so straightforward. The problem is that when we focus our attention on emerging digital technologies, and we reflect on their potential significance to philosophical debates regarding extended cognition and extended knowledge, a number of worrying tensions come to light. These tensions do not, I suggest, affect the legitimacy of claims made about the nature of knowledge or the potentially extended character of human cognition—there is, for example, no reason to doubt the basic possibility of Web-extended cognizers. Nor do the tensions have any impact on the apparent (conceptual) compatibility of active externalism and virtue reliabilism when it comes to the possibility of extended knowers. Rather, the tensions concern the way in which the properties of emerging digital technologies relate to the positive epistemic standing of a technologically extended cognitive agent.

One of the main tensions relates to the fact that many of the properties that seem to make a technology apt for cognitive incorporation are also ones that seem to threaten or undermine the epistemic standing of an extended cognizer. In order to make sense of this claim, let us focus our attention, for the moment, on the personalization criterion. Personalization, recall, involves an external resource being tailored to the agent who uses it. Such tailoring plausibly works to improve the interaction between the agent and the resource, and it also serves to enhance the accessibility of information that is provided by the resource. Earlier in this chapter, I outlined a number of ways in which emerging digital technologies could satisfy the personalization criterion. In addition to the active configuration of the device, I outlined a variety of ways in which the information-processing profile of a technological device, or its presentational capabilities, could be aligned to fit with an agent’s idiosyncratic modes of use and areas of epistemic interest. But now see how this capacity for personalization leads to an apparent tension. For the capacity of technologies to adjust their modes of operation in respect of our usage patterns has emerged as a major source of epistemological concern (Miller and Record 2013; Simpson 2012). Consider, for example, the case of personalized search. Many modern search engines rely on user profiling techniques to tailor search results to the information retrieval requirements of specific end users.

In one sense, this form of automatic personalization enhances the ability of the search engine to align its mode of operation with the epistemic needs and interests of a particular Web user. However, personalized search mechanisms have also been linked to the formation of what are called “filter bubbles” (Pariser 2011), and these have been seen to limit the justificatory status of a user’s beliefs and threaten their overall epistemic standing (see Miller and Record 2013). Inasmuch as such claims are true, it seems that we confront a situation in which a technology is satisfying at least one of the criteria for cognitive extension (i.e., personalization) while simultaneously posing a threat to our status as epistemic agents. The lesson, it seems, is that the properties that make a technology apt for cognitive incorporation are not always the ones that yield an epistemically desirable state of affairs. In fact, the very same features that make a technology apt for cognitive incorporation may also (on occasion) result in states of affairs that undermine the epistemic credentials of the technology-wielding agent.

One response to this sort of worry is to suggest that when it comes to personalization what really matters is that an agent has some form of control over the personalization process. Alternatively, we might seek to challenge the claim that personalized search really does pose a threat to our epistemic well-being (see Smart and Shadbolt In press). Unfortunately, however, the concern regarding personalization is merely one example of a more general worry. Recall, for example, that the accessibility criterion mandates that the information provided by a resource should be easily accessible—the more accessible the information is, the more suitable the resource seems to be as a constituent element in an extended cognitive process. This criterion implies that a resource should be easy to use, or at least it should have the kinds of properties that enable it to (eventually) become transparent in use. Central to these notions of accessibility and transparency, I suggest, is the idea of fluency, which is defined as the “subjective experience of ease or difficulty with which we are able to process information” (Oppenheimer 2008, 237). It seems, therefore, that engineers with an interest in cognitive extension should seek to target a set of properties that yield fluent forms of interaction between a human user and a technological resource. Now the problem: fluency is something that influences our epistemic judgments regarding the truth status of processed information (Reber and Schwarz 1999; Reber and Unkelbach 2010). In particular, the more fluently we are able to process information, the more likely we are to fall foul of a truth bias, i.e., to overestimate the truth status of processed information. It should be clear, then, that fluency is pulling us in different directions with regard to extended cognition and extended knowledge. In particular, it seems that if individuals are provided with quick and easy access to information as a result of some form of bio-technological bonding, they may be more inclined to view the information as true. As was the case for personalization, the main point here is simply that a property of a technology that seems to make it a better candidate for cognitive incorporation is also one that raises concerns about its impact on our epistemic profile. There is, in other words, a technology-induced tension between our status as extended cognizers and extended knowers.

There are other, perhaps more subtle, ways in which this sort of tension comes to light. Recall that in discussing linked data, I sought to highlight the value of linked data formats in improving our access to Web-accessible information. The embedding of information in the context of conventional Web pages was, I suggested, something that reduced the accessibility of information, but the accessibility of information could be enhanced by resorting to the representational strategies adopted by the linked data community. As a result of the kind of enhanced accessibility that linked data formats provide, we thus seem to have a technological and informational infrastructure that is more suitably poised to support the emergence of Web-extended minds. When seen from an epistemological perspective, however, the transition to linked data formats presents us with a problem. The problem is that by removing information from its traditional location within a Web page, we remove many of the contextual cues and affordances that might otherwise have played a productive role in the epistemic evaluation of the target content. A variety of human user studies attest to the value of such cues in the evaluation of Web-based information (Metzger 2007). Fogg et al. (2003), for example, report that when users are evaluating online content, their credibility judgments are likely to be influenced by factors relating to the visual design of the website. In fact, they are more likely to be influenced by the visual design of a website than they are by its actual content! Contextual cues thus seem to play an important role in supporting credibility judgments, and they are therefore, at least in principle, an important source of information regarding the reliability of online content. The concern is that by moving to linked data formats and thereby enhancing the accessibility of online information, we also remove the kind of informational cues that support the adaptive selection, monitoring, and adjustment of belief-forming processes, as well as the selective endorsement of belief-relevant information. The basic point, again, is that by introducing technologies that speak to the various criteria for cognitive extension, we inadvertently create a situation in which the epistemic standing of an extended agent seems to be called into question.

In closing this section, I want to highlight a further concern regarding emerging digital technologies and extended knowledge. This is a concern that arises out of a consideration of the way in which new technologies are transforming our cognitive ecology (see Hutchins 2010). In order to better understand this concern, consider that in the introduction to this chapter I mentioned that the rate of technological innovation and change is greater than it has ever been before. In subsequent sections, I cast technological change and innovation in a largely positive light relative to our status as extended cognitive agents. Emerging digital technologies, I suggested, speak to many of the conditions that guide our intuitions as to when we confront a genuine case of cognitive extension. In view of all this, active externalists should be perfectly happy to embrace a form of “technological progressivism” regarding technology development.

But now consider the situation from the perspective of virtue reliabilism. According to the virtue reliabilist, abilities play a key role in establishing positive epistemic status.

Abilities, however, are the sorts of things that typically take time to acquire.¹⁴ They are also the sorts of things that are affected by changes in the environment. If one changes the normal ecological context in which abilities develop, then there is at least the possibility that the processes reflecting the exercise of that ability will be undermined. Both the development and the expression of abilities thus depend on a certain degree of ecological or environmental stability, and given that rapid rates of technological innovation and change seem to threaten the stability of our cognitive ecologies, it is far from clear that an advocate of virtue reliabilism can see emerging technologies in quite the same way as does the active externalist. In fact, we might say that the virtue reliabilist is much more comfortable embracing a form of “technological conservatism”—countenancing an approach that calls for more gradual (or more carefully managed) forms of technological change. On the surface, therefore, it might seem as if virtue reliabilism and active externalism are at least compatible, if not entirely natural, bedfellows. This compatibility, however, seems to be at risk once we consider the effect of technological change on our epistemically relevant cognitive abilities.

One response to this apparent tension is to focus on the kinds of abilities that I have suggested are relevant to fixing the truth of an agent’s beliefs—the sub-personal metacognitive mechanisms of selection, monitoring, and endorsement. These abilities, it seems, are unlikely to be affected in any dramatic fashion by the changing nature of our technological environment. Given that, in most cases, the abilities in question supervene on processes that are contained within the agent’s biologically based organismic boundaries, it seems that they may be largely unaffected by new technologies. It is in this respect that the non-extended nature of the processes perhaps counts as a blessing for virtue reliabilism. The problem, however, is that it is far from clear that these metacognitive mechanisms can function in a way that preserves or enhances an agent’s ability to track the truth in the face of sudden and rapid shifts in the technological terrain. Such shifts come in a variety of forms, but arguably the most epistemically pernicious kind of change occurs when the operation of a technology is subject to covert modification. Software upgrades that affect the behind-the-scenes behavior of online services and locally installed applications are one example here. The effect of such ‘invisible’ changes is to potentially alter the functional profile of some bio-external resource in a way that introduces a hidden vulnerability,¹⁵ or which modifies the reliability of cues that would otherwise indicate the contingent reliability of first-order belief-forming processes. As is evident from cases in the epistemological literature, most notably the Barn Facade case (see Pritchard 2009), undetected changes in an agent’s environment can wreak havoc in terms of the tenability of knowledge attributions.

¹⁴ This is an assumption that rests on a particular understanding of what abilities are. Palermos (2013), for example, comments that abilities should be seen as “... reliable (cognitive) processes, [that] are acquired over a long period of time through constant interaction of the agent with the environment wherein he is typically situated” (1196).

¹⁵ Clark (2015) also alludes to a “hidden vulnerability” in accounting for our reluctance to credit agents with knowledge.

3. Conclusion

In this chapter, I have attempted to advance our understanding of the cognitive and epistemic implications of emerging digital technologies when considered from the perspective of both active externalism and virtue reliabilism. The value of emerging digital technologies is that they highlight issues that might not have been revealed if we had limited ourselves to the rather mundane and technologically low-grade examples favored by the philosophical community. In general, emerging digital technologies possess features that are broadly consistent with the sort of criteria that have been used to evaluate putative cases of cognitive extension. In fact, if anything, the features of emerging digital technologies seem to make them better candidates for cognitive incorporation, as compared to the kind of artifacts that typically animate philosophical discussions in this area.

When it comes to virtue-theoretic conceptions of knowledge, we saw that virtue reliabilism is largely compatible with active externalism. Importantly, however, active externalism does not seem to warrant a retreat from the more robust (or at least strong cognitive agency) variants of virtue reliabilism. This is because explanations as to why an agent believes the truth will often need to make reference to processes that determine the adaptive configuration and monitoring of cognitively relevant circuits that shape the course of intelligent behavior. Such processes are usefully treated as examples of metacognition. They influence the choice between inner and outer sources of information, the shape of belief-forming processes, and the control that information has over behavioral output. Given that these processes are typically realized by neural mechanisms, there does not seem to be any reason to cater for a putative distribution of credit when it comes to understanding the epistemic status of technologically extended agents. Irrespective of the nature of the first-order belief-forming processes (e.g., whether they are extended or non-extended), the abilities that seem to determine whether or not an agent believes the truth are ones that remain within the head of the agent. Knowledgeable agents, I suggest, are deserving of *primary* credit (or blame) for the beliefs they embrace.

Relative to this view, it seems that the impact of emerging digital technologies on our epistemic profile should be relatively benign. In fact, to the extent that such technologies open the door to cognitively potent forms of bio-technological merger that expand the number of beliefs we have, we might expect emerging technologies to enhance our epistemic profile (Bjerring and Pedersen 2014; Ludwig 2015). I have attempted to sound a note of caution here. In particular, a consideration of both the properties of new technologies, as well as the way in which such technologies transform the nature of the cognitive ecologies in which our beliefs are formed, seems to give rise to a number of apparent tensions between active externalism and virtue reliabilism. All of this should serve to remind us that the cognitive and epistemic benefits of a new technology are not a given; good design entails a commitment to understanding the way in which a technology will change the nature of both ourselves and the societies in which we live. When it comes

to the design and development of new technologies, we should therefore reflect on the kinds of technologies that best serve our cognitive, epistemic, and social interests. The debate concerning extended cognition and extended knowledge has much to offer here. By situating epistemological concerns at the heart of our consideration of technologically extended cognitive systems, we are in a better position to understand not just what it is that makes a technology apt for cognitive incorporation but also what it is that makes a technology worth bonding with.

References

- Anderson, M. L., M. J. Richardson, and A. Chemero. (2012). "Eroding the Boundaries of Cognition: Implications of Embodiment." *Topics in Cognitive Science* 4 (4): 717–30.
- Berners-Lee, T., J. Hendler, and O. Lassila. (2001). "The Semantic Web." *Scientific American* 284 (4): 34–43.
- Bizer, C. (2009). "The emerging web of linked data." *IEEE Intelligent Systems* 24 (5): 87–92.
- Bizer, C., J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. (2009). "DBpedia-A crystallization point for the Web of Data." *Web Semantics: Science, Services and Agents on the World Wide Web* 7 (3): 154–65.
- Bjerring, J. C., and N. J. L. L. Pedersen. (2014). "All the (many, many) things we know: Extended knowledge." *Philosophical Issues* 24 (1): 24–38.
- Carr, N. (2010). *The Shallows: How the Internet Is Changing the Way We Think, Read and Remember*. London: Atlantic Books.
- Clark, A. (1997). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA: MIT Press.
- Clark, A. (2005). "Intrinsic content, active memory and the extended mind." *Analysis* 65 (285): 1–11.
- Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York: Oxford University Press.
- Clark, A. (2010a). "Coupling, Constitution and the Cognitive Kind: A Reply to Adam and Aizawa." In R. Menary (ed.). *The Extended Mind*, 81–99. Cambridge, MA: MIT Press.
- Clark, A. (2010b). "Memento's revenge: The extended mind, extended." In R. Menary (ed.). *The Extended Mind*, 43–66. Cambridge, MA: MIT Press.
- Clark, A. (2013). "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science." *Behavioral and Brain Sciences* 36 (3): 181–253.
- Clark, A. (2015). "What 'Extended Me' knows." *Synthese* 192 (11): 3757–75.
- Clark, A. (2016). *Surfing Uncertainty: Prediction, Action and the Embodied Mind*. New York: Oxford University Press.
- Clark, A., and D. Chalmers. (1998). "The Extended Mind." *Analysis* 58 (1): 7–19.
- Clowes, R. (2015). "Thinking in the Cloud: The Cognitive Incorporation of Cloud-Based Technology." *Philosophy & Technology* 28 (2): 261–96.
- Fogg, B., C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, and E. R. Tauber. (2003). *How do users evaluate the credibility of Web sites? A study with over 2,500 participants*. San Francisco, CA: Conference on Designing for User Experiences, New York: ACM.

- Goldman, A. I. (2011). "A Guide to Social Epistemology." In A. I. Goldman and D. Whitcomb (eds.). *Social Epistemology: Essential Readings*, 11–37. Oxford and New York: Oxford University Press.
- Greco, J. (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
- Hassabis, D., and E. A. Maguire. (2007). "Deconstructing episodic memory with construction." *Trends in Cognitive Sciences* 11 (7): 299–306.
- Heath, T., and C. Bizer. (2011). "Linked Data: Evolving the Web into a Global Data Space." *Synthesis Lectures on the Semantic Web: Theory and Technology* 1 (1): 1–136.
- Hutchins, E. (2010). "Cognitive ecology." *Topics in Cognitive Science* 2 (4): 705–15.
- Kelp, C. (2013). "Extended cognition and robust virtue epistemology." *Erkenntnis* 78 (2): 245–52.
- Kelp, C. (2014). "Epistemology extended." *Philosophical Issues* 24 (1): 230–52.
- Koriat, A., and R. Nussinson. (2009). "Attributing study effort to data-driven and goal-driven effects: Implications for metacognitive judgments." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35 (5): 1338–43.
- Ludwig, D. (2015). "Extended cognition and the explosion of knowledge." *Philosophical Psychology* 28 (3): 355–68.
- Matthen, M. (2010). "Is memory preservation?" *Philosophical Studies* 148 (1): 3–14.
- Metzger, M. J. (2007). "Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research." *Journal of the American Society for Information Science and Technology* 58 (13): 2078–91.
- Meyrand, P., J. Simmers, and M. Moulins. (1994). "Dynamic construction of a neural network from multiple pattern generators in the lobster stomatogastric nervous system." *The Journal of Neuroscience* 14 (2): 630–44.
- Michaelian, K. (2012a). "Is external memory memory? Biological memory and extended mind." *Consciousness and Cognition* 21 (3): 1154–65.
- Michaelian, K. (2012b). "Metacognition and endorsement." *Mind & Language* 27 (3): 284–307.
- Michaelian, K. (2014). "JFGI: From distributed cognition to distributed reliabilism." *Philosophical Issues* 24 (1): 314–46.
- Michaelian, K., and S. Arango-Muñoz. (2014). "Epistemic feelings, epistemic emotions: Review and introduction to the focus section." *Philosophical Inquiries* 2 (1): 97–122.
- Miller, B., and I. Record. (2013). "Justified belief in a digital age: On the epistemic implications of secret Internet technologies." *Episteme* 10 (02): 117–34.
- Oppenheimer, D. M. (2008). "The secret life of fluency." *Trends in Cognitive Sciences* 12 (6): 237–41.
- Oren, E. (2005). *SemperWiki: A semantic personal Wiki*. Galway, Ireland: International Conference on Semantic Desktop Workshop. Aachen, Germany: CEUR-WS.org.
- Palermos, S. O. (2011). "Belief-forming processes, extended." *Review of Philosophy and Psychology* 2 (4): 741–65.
- Palermos, S. O. (2013). "Could Reliability Naturally Imply Safety?" *European Journal of Philosophy* 23 (4): 1192–208.
- Palermos, S. O. (2014a). "Knowledge and cognitive integration." *Synthese* 191 (8): 1931–51.
- Palermos, S. O. (2014b). "Loops, Constitution, and Cognitive Extension." *Cognitive Systems Research* 27: 25–41.

- Palermos, S. O. (2015). "Active externalism, virtue reliabilism and scientific knowledge." *Synthese* 192 (9): 2955–86.
- Palermos, O., and D. Pritchard. (2013). "Extended Knowledge and Social Epistemology." *Social Epistemology Review and Reply Collective* 2 (8): 105–20.
- Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin.
- Preston, J. (2010). "Extended Mind, the Concept of Belief, and Epistemic Credit." In R. Menary (ed.), *The Extended Mind*, 355–69. Cambridge, MA: MIT Press.
- Pritchard, D. (2009). *Knowledge*. Basingstoke: Palgrave Macmillan.
- Pritchard, D. (2010). "Cognitive ability and the extended cognition thesis." *Synthese* 175 (1): 133–51.
- Pritchard, D. (2012). "Anti-luck virtue epistemology." *The Journal of Philosophy* 109 (3): 247–79.
- Proust, J. (2014). "Epistemic action, extended knowledge, and metacognition." *Philosophical Issues* 24 (1): 364–92.
- Reber, R., and N. Schwarz. (1999). "Effects of perceptual fluency on judgments of truth." *Consciousness and Cognition* 8 (3): 338–42.
- Reber, R., and C. Unkelbach. (2010). "The epistemic status of processing fluency as source for judgments of truth." *Review of Philosophy and Psychology* 1 (4): 563–81.
- Reitter, D., and C. Lebiere. (2012). *Social Cognition: Memory Decay and Adaptive Information Filtering for Robust Information Maintenance*. Toronto: Canada: 26th AAAI Conference on Artificial Intelligence. Palo Alto, California: AAAI Press.
- Riggs, W. (2007). "Why epistemologists are so down on their luck." *Synthese* 158 (3): 329–44.
- Schacter, D. L., and D. R. Addis. (2007). "The cognitive neuroscience of constructive memory: Remembering the past and imagining the future." *Philosophical Transactions of the Royal Society B: Biological Sciences* 362 (1481): 773–86.
- schraefel, m. c., D. A. Smith, A. Owens, A. Russell, C. Harris, and M. L. Wilson. (2005). *The evolving mSpace platform: Leveraging the Semantic Web on the trail of the memex*. Salzburg, Austria: 16th Conference on Hypertext and Hypermedia. New York: ACM.
- Simpson, T. W. (2012). "Evaluating Google as an epistemic tool." *Metaphilosophy* 43 (4): 426–45.
- Smart, P. R. (2012). "The Web-Extended Mind." *Metaphilosophy* 43 (4): 426–45.
- Smart, P. R. (2014). "Embodiment, Cognition and the World Wide Web." In L. A. Shapiro (ed.), *The Routledge Handbook of Embodied Cognition*, 326–34. New York: Routledge.
- Smart, P. R., P. C. Engelbrecht, D. Braines, M. Strub, and C. Giammanco. (2010). "The Network-Extended Mind." In D. Verma (ed.), *Network Science for Military Coalition Operations: Information Extraction and Interaction*, 191–236. Hershey, PA: IGI Global.
- Smart, P. R., and N. R. Shadbolt. (2014). "Social Machines." In M. Khosrow-Pour (ed.), *Encyclopedia of Information Science and Technology*, 6855–62. Hershey, PA: IGI Global.
- Smart, P. R., and N. R. Shadbolt. (In press). "The World Wide Web." In J. Chase and D. Coady (eds.), *The Routledge Handbook of Applied Epistemology*. New York: Routledge.
- Smart, P. R., K. Sycara, and Y. Tang. (2014). *Using Cognitive Architectures to Study Issues in Team Cognition in a Complex Task Environment*. Baltimore, MD: SPIE Defense, Security, and Sensing: Next Generation Analyst II., Bellingham, Washington: SPIE.
- Sterelny, K. (2010). "Minds: Extended or scaffolded?" *Phenomenology and the Cognitive Sciences* 9 (4): 465–81.

- Van Kleek, M., D. A. Smith, N. Shadbolt, and m. c. schraefel. (2012). *A decentralized architecture for consolidating personal information ecosystems: The WebBox*. Seattle, WA: Personal Information Management.
- Weiskopf, D. A. (2008). "Patrolling the Mind's Boundaries." *Erkenntnis* 68 (2): 265–76.
- Wikforss, Å. (2014). "Extended belief and extended knowledge." *Philosophical Issues* 24 (1): 460–81.
- Wilson, R. A., and A. Clark. (2009). "Situated Cognition: Letting Nature Take its Course." In P. Robbins and M. Ayded (eds.). *The Cambridge Handbook of Situated Cognition*, 55–77. Cambridge: Cambridge University Press.