Emerging Digital Technologies: Implications for Extended Conceptions of Cognition and Knowledge
Paul Smart, Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ, UK

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, for example, the World Wide Web has transformed the way we shop, date, socialize and undertake scientific endeavours. 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. 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 concerning the extent to which technologies are, on occasion, incorporated into the fabric of society (e.g., Smart & 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 the Web could have been abandoned and society would probably not have been affected in any major way. Fast forward 30 years, however, and it seems much harder to decouple the Web from society. The Web now plays a role in society that is akin to the role 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 up our society. 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 has 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, in this case, become inextricably linked, and both have been transformed as a result of the socio-technological merger.

1 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.
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. At the heart of this debate we encounter 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 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 kind 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 current chapter. The current chapter also aims to examine the nature of the relationship between emerging digital technologies and philosophical conceptions of extended knowledge (Palermos & 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.

Part I: Extended Cognition

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 & Chalmers, 1998). This idea is sometimes presented as a thesis regarding 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 which he uses to store important pieces of information. Otto is so reliant on the notebook and so accustomed to using it that he carries the notebook with him wherever he goes and accesses the notebook fluently and automatically whenever he needs to do so. Having thus set the stage, Clark and Chalmers (1998) 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 street, and then walks to the museum. It is clear that in making this episode of behaviour intelligible (or psychologically transparent) to us Inga must have desired to go to the museum, and it is clear that she walked to 53rd street because that is where she believed the museum was located. Obviously, Inga did not believe that the museum was on 53rd street in an occurrent sense (i.e., she has not spent her entire
life consciously thinking about the museum’s location); rather, she entertained the belief in a dispositional sense. Inga’s belief, like perhaps many of her beliefs, was sitting in memory, waiting to be accessed as and when needed.

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

Clark and Chalmers (1998) argue that the Otto thought experiment establishes the case for a form of externalism about Otto’s states of dispositional believing. The notebook, they argue, plays a role that is functionally akin to the role played by Inga’s onboard bio-memory. If this is indeed the case, then it seems to make sense to see the notebook as part of the material supervenience base for some of Otto’s mental states, specifically his states of dispositional belief (such as those involving museum-related facts). The main point of the argument is thus to establish a (potential) role for non-biological artefacts 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.

**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) rightly notes, “There would be little value in an analysis that credited me with knowing all the facts in the Encyclopaedia Britannica just because I paid the monthly instalments and found space for it in my garage” (p. 217). The nature of the relationship between the human agent and the encyclopaedia, in this case, seems to be such that talk of the encyclopaedia as counting as part of the realization base for one’s states of dispositional belief seems utterly inappropriate.

In order to guard against the problem of cognitive bloat, a number of criteria have been proposed to limit candidate cases of cognitive extension. 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, p. 46).
- **Accessibility**: the “information contained in the resource should be easily accessible as and when required” (Clark, 2010b, p. 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, p. 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, p. 274).

• **Continuous Reciprocal Causation:** “continuous mutual interactions between the organismic agent and the artefact...[are] both necessary and sufficient for cognitive extension” (Palermos, 2014b, p. 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 kind 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 technologies might enable them to support current or near-future forms of cognitive extension.

**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 attracted relatively little 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 technologies in enhancing (or perhaps undermining) our cognitive capabilities at both the individual and collective level.

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 broad range of cognitive, social and scientific endeavours, 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 profile? Could the Web herald a new era in which our body of knowledge and beliefs is limited only by the sort of informational contact we have with the online digital environment?

One finds a positive response to these sort 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” (p. 356). Similar views are expressed by Bjerring and Pedersen (2014). They 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” (p. 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 kind of criteria that have been proposed to help us discriminate genuine cases cognitive extension from the more ersatz varieties (see above). Recall that the motivation for specifying such criteria in the first place was to avoid precisely the kinds of cognitive bloat that both Ludwig (2015) and, to some extent, Bjerring and Pedersen (2014) warmly embrace. 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) thus seem 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 relating to availability, accessibility and trust. In more recent work, Clark (2010b) suggests that mobile access to Google should not count 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. However, I too voiced a number of concerns about the extent to which the nature of our informational contact with the Web could satisfy the criteria for cognitive extension. A particular concern, I suggested, was 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 are typically the case when we attempt to access information from Wikipedia. The situation is, however, often worse when we attempt to access information from institutional websites. 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.

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 situated at the top of the article and is pretty much the first fact we encounter. Darwins’ 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 it, I suggest that tying accessibility to the structural layout of a dynamically structured information resource is a risky business. Particular facts may, by virtue of their 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 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 really that a particular form of Web-based interaction – one characterized by the use of conventional Web browser technologies and document-centred 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 Web access tell us everything about the nature of our contact with the online environment. Crucially, the Web is a system in a state of more-or-less constant technological flux. Although it might be useful to think of ‘the Web’ as identifying a particular kind of system, the extensional projection of this term is, in fact, likely to subsume an array of highly disparate systems, each varying with respect to the kind of interactivity and connectivity they provide to their human user base. The Web, as such, is 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 & Pedersen, 2014; Ludwig, 2015).

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 the 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.

Linked Data and the Semantic Web
When most of us think of ‘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 of the Web encourages us to think of the Web as a globally-distributed network of documents connected by hyperlinks. Call this the document-centred Web, or the Web of Documents.

In addition to the document-centred Web, another vision of the Web has surfaced in recent years. This is the vision of the data-centred Web, or the Web of Data (Bizer, 2009; Heath & Bizer, 2011). The basic idea, here, is that instead of seeing 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 extend the reach of a broad range of social, scientific and political activities.

The precursor to research on linked data is an initiative that goes under 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 is sometimes treated independently of linked data, it is important to note that both the linked data and Semantic Web 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 ontologies 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 inference and reasoning.

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, multiple points of interest here, not of all of which I am able to elaborate on in the current chapter. 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, bedevilled the information retrieval efforts of human subjects in the context of the document-centred 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 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 way in which data can be used to drive the creation of a potentially unlimited variety of informational artefacts, 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 kind of device (e.g., laptop versus Smartphone) that is used to access it. Linked data

2 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.
assets can obviously support this kind of device-related flexibility; they are also, however, 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 significance when we consider 1) the broad array of Web-enabled devices that are now starting to emerge, and 2) the many and varied task contexts in which a particular online data asset might be called into productive cognitive service.

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, often one that is 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 informational encodings in a way that is broadly compatible with the possibility of Web-extended minds. There is, however, something of particular interest here. Note that the typical wording of the availability criterion (see above) 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 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 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 very 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 located 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 specific devices and particular 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 might 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 we typically encounter in offline contexts. One of these features is ‘autonomy’, which refers to the way in which cloud-based information is amenable to a variety of 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 the use of linked data formats and semantically-enriched representations (see above). 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 individuals 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 what have been referred to as their ‘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 sort of systems are emblematic of the way in which the Web can be used to support access to highly personalized bodies of information and data.

**Wearable Technology**

Wearable technologies epitomize the current trend towards greater mobility and portability in the design of digital technologies. These technologies now come in a bewildering array of forms, such as smart watches, EEG headsets, activity trackers, smart clothes, digital jewellery and so on. Given that these technologies are, by definition, designed to be worn by a user, it should come as no surprise that wearable technologies speak directly 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 is thus something that he needs to carry around with him, and it is therefore liable to interfere with many of his other 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) enables computational processing to be sensitized to various features of the context in which it 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 contextualise the search effort, yielding a set of geographically filtered search results that are likely to be highly germane to my current needs and interests. Such forms of contextualized processing are likely to become increasingly sophisticated as we witness the emergence of new sensor-enabled wearable devices. Future capabilities could include the contextualization of information retrieval processes to a range of social, behavioural 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 & 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 & Nussinson, 2009). By being able to detect such physiological responses, future technologies might adaptively modify their modes of operation to support human end-users with regard to a variety of epistemically-related activities.

**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, I suggest we can see these technologies as creating virtual variants of the designer environments discussed by Clark (2008, p. 59). Such ‘virtual designer environments’ provide a number of opportunities to reshape the real-world environment in ways that extend the reach of our own cognitive capabilities. By extending the real-world with highly malleable digital representations, the profile of our own cognitive endeavours 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 do enable 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. 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 & Clark, 2009), then it seems entirely appropriate to see the task involving the hologrammatic representations as a form of extended cognizing as well.

Thus far, I have said little 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 (see above). The specific claim is that there should be some form of mutual influence and interactive engagement between the agent and the resource, such that the agent influences the state of the resource, which then exerts a feedback influence on the agent, and so on. The general idea is thus that extended cognitive
systems are ones in which we encounter reciprocal forms of information flow and influence between the constituent elements of some larger cognitive system (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 levied 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. Invariably such changes will influence the kind of actions performed by the human agent. This sort of interactive engagement between device and user is particularly apparent in the case of hologrammatically-extended cognition. Here, the actions of the user leads 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, however, the user is exerting a form of indirect influence on the device, one which 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 artefact.

**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 the device are likely to reflect the specific informational needs and interests of the user, as well as the kind of tasks in which they are 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, there some current 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. The most high profile example of this sort of development effort is Google’s Project Ara³. 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.

A third form of personalization centres on the use of user profiling techniques, which are sometimes used in to monitor the behaviour of (e.g.) Web users in order to make predictions about their tastes and preferences. Perhaps the most well-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 into 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.

**Web-Extended Minds, Revisited**

**Technologically-Extended Otto**

The foregoing 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 the same extent as the usual philosophical cases, most of which rely on rather primitive technological resources. 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, in terms of guiding thought and action, seem 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 levied 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 Street. The failure to exhibit such forms of updating 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 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 towards 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 an additional form of updating. Importantly, such updates (providing the various sources of information are reliable) can be expected to yield informational encodings that reliably 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.

**Trusting the Web: Endorsement and Selection**

No doubt the astute reader will be keen to point out number of problems with the discussion thus far. 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 or assertion 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 seem 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 we rely on 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 possess a fair number of false beliefs. In fact, we might go as far 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 fulfil the criteria for cognitive extension and knowledge attribution) starts to seem a little remote.

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 things even 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 our onboard biological memory systems. 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 go 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 the presence of 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 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<sub>STRONG</sub>) is
that externally-situated information should not be subject to any form of evaluative assessment. The
second reading (AE<sub>WEAK</sub>) 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 referencing 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<sub>WEAK</sub>). It should be clear that, at least from an epistemic perspective, AE<sub>WEAK</sub> is also the
preferred reading. If we were obliged to commit to AE<sub>STRONG</sub>, 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<sub>STRONG</sub> in favour of AE<sub>WEAK</sub>. Now, it seems, we can
allow for the possibility of some form of evaluative assessment. In fact, one reason to accept AE<sub>WEAK</sub>
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 &
Maguire, 2007; Matthen, 2010; Schacter & 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, here, that something has gone awry, 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 towards the
stronger reading of the trust criterion, i.e., AE<sub>STRONG</sub>. If the extended mind theorist is indeed
committed to AE<sub>STRONG</sub>, and if Michaelian is right about endorsement in the case of bio-memory,
then there is clearly an issue that is in need of further thought and reflection. However, what if the
extended mind theorist is only committed to the weaker reading of the trust criterion. In this case, much of the sting seems to be taken out of Michaelian’s (2012a) critique⁴.

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\textsubscript{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 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 seems somewhat redundant – it cannot enhance the reliability of a process that is already 100% 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 is 100% accurate, the second order process still needs to be accurate in terms of its reliability estimates concerning the first-order process (i.e., the second-order process itself needs to be reliable). Imagine a situation where the first-order process is 100% accurate, but the second-order process accurately estimates the reliability of the first-order process only 50% 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 process⁵.

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, p. 321). The selection problem is thus a problem of selecting an appropriate cognitive strategy in a given context. In the case of extended cognition, the primary problem confronting the agent is one that 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 select 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

⁴ If Michaelian (2012a) is wrong about biologically-based endorsement, of course, then it does not seem to 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.

⁵ 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.
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 the relevant resources, the assembly of the distributed [i.e., extended] system, and the endorsement of information produced by the system” (pp. 320-321). The significance of such claims will become increasingly apparent when we examine the role played by cognitive agency in virtue reliabilistic approaches to knowledge attribution (see below). Also note that by linking the selection process to the relative reliability (i.e., accuracy) of different information sources, we might go some way towards addressing concerns about the original (trust-related) endorsement problem. In fact, inasmuch as Jesse is selecting processes on their basis of their tendency to yield reliable information, then we might wonder whether there is a need for him to engage in the kinds of epistemically-hygienic practice associated with AESTRONG; i.e., the tendency to subject information to conscious scrutiny and deliberate evaluation. Perhaps all Jesse really needs to do in this situation is be responsive to epistemically-relevant factors that indicate errors concerning 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 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 respect to his 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 of 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 target 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 process 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.

**Metacognition and the Network-Extended Mind**

Michaelian (2012b, 2014) suggests that we should see endorsement and selection as examples of metacognitive processes; i.e., 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 concerning the adaptive construction and configuration of cognitively-relevant information processing circuits. This idea is apparent in the 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, p. 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 the process of selection results in the creation of a dynamically configured (or dynamically assembled) network that alters the information processing profile of the larger systemic 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 his hands are 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 ‘effective’ connectivity of socio-technical ensembles can be dynamically modified in order to meliorate collective cognitive processing in distributed or collective cognitive systems (e.g., Smart et al., 2014). One study that is of 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 behaviour can optimize the spread of information through a communication network” (p. 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 cognitive systems, such as collective doxastic agents (Palermos & 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 (in press). Clark (in press) 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 those 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 (in press) analysis enables us to 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) 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
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 behaviours 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).

Part II: Extended Knowledge

Virtue Reliabilism

One of the driving forces behind recent work in epistemology is the intuition that an agent’s true beliefs should count as knowledge if the beliefs arise as the result of the exercise of an agent’s cognitive abilities. This intuition, commonly referred to as the ability intuition (Pritchard, 2009), lies at the heart of popular theory of knowledge that goes under the name of virtue reliabilism. A key feature of virtue reliabilistic accounts is the emphasis they place on cognitive abilities in terms of understanding why it is that an agent believes the truth. 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 the constitute the agent’s cognitive character. Knowledge is thus seem as a form of success that results from the exercise of ability.

There are actually 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 an agent believes the truth.

An alternative to robust virtue reliabilism comes in the form of the ‘weak cognitive agency’ (COGA\textsubscript{WEAK}) view expressed by Pritchard (2010):

\textbf{COGA\textsubscript{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\textsubscript{WEAK} differs from robust virtue reliabilism in two ways. Firstly, the commitment to the biconditional in the case of COGA\textsubscript{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\textsubscript{WEAK} relates to the degree to which cognitive success is creditable to cognitive agency. In the case of robust virtue epistemology we encountered the notion of primary creditability. COGA\textsubscript{WEAK} reduces the extent to which the truth of a belief is creditable to agential abilities by invoking the notion of *significant* creditability. It is this weakening of the credibility constraint that is deemed to make COGA\textsubscript{WEAK} particularly well-suited 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 the agent believes, with some of the credit being attributable to non-biological elements. Inasmuch as we accept this claim, then it seems that an emphasis on primary creditability is likely to cause problems in terms of accounting for knowledge in the case of extended agents. In particular, it seems that, in some cases, what an agent believes may have more to do with the properties of some bi-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\textsubscript{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 concept of extended knowledge.

\textbf{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 way in which technologically-advanced forms of cognitive extension might contribute to positive epistemic standing qua virtue reliabilistic accounts. 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 notions of extended cognitive processes, cognitive ability and epistemically-relevant credit attributions.

\footnote{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” (p. 8). See Preston (2010) for a critical analysis of this claim.}
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 in its physical realization. Clearly, this view sits very comfortably alongside claims concerning 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 principal 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 at hand, it seems relatively clear what the nature of those conditions should be. In particular, what needs 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 that thus seem to be of interest are 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 clearly 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 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 thing we could say here is that it is not so much the content of the belief state 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. This seems to serve as a useful 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 to the problem might be to claim 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 seemingly contribute to the formation of an agent’s dispositional beliefs and focusing our attention solely on the run-time processes, we seem to 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 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, qua 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 behaviour) 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. 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.

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 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” (p. 88). The idea, then, seems to be 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

---

9 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 preconfigured 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 with regard to the truth of his beliefs.

10 Note that this sort of problem only comes about as a result of considering the role that belief-forming processes play in securing positive epistemic status in the context of virtue reliabilistic conceptions of knowledge. This is not necessarily a dilemma faced by the extended mind theorist who focuses solely in extended states of dispositional belief.
before the cognitive circuit, incorporating the external resource, is actually established. 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.

**Extended Belief-Forming Processes**

Given the difficulty in tying the notion of extended knowledge to extended realization bases for 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 the matters in the case of extended knowledge is that an agent possesses knowledge in virtue of having formed a belief on the basis on 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 seems to commit 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, with this new view of extended knowledge in place, we are now in a position to ask 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. Both Otto++ and Jesse can 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 explosion in knowledge that is envisaged by 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 their deployment of extended cognitive processes is irrelevant to explanations as to why this explosion could occur. Rather, the problem is that it is far from clear that it is the extended process itself that is doing all 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” (p. 745). Similarly, Greco (2010) claims that “in 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” (p. 75).
The question, therefore, is to what extent do Otto++ and Jesse participate in, or deploy, extended cognitive processes that explain their success in believing the truth? Is it the extended cognitive processes that determines the truth status of their beliefs...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 endeavours 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 degree 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...es he believes to be correct. Here, the character are of ascribing knowledge to Jesse, selection of a particular explanatory account lines up with the facts (i.e., whether it tracks the truth) seems to determining what it is that Jesse believes, i.e., extended cognitive process, truth playing an explanatorily significant role when it comes to accounting for why Jesse believes he should not believe relative these source resources...pursue a fair degree of epistemic discernment and cognitive ability. But what is it that...nn, there. The monitoring of its informational deliverances...ut this belief lines up with the facts (i.e., whether it tracks the truth) seems to require more than just a focus on 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 mechanisms involved in the selection of a particular process, the monitoring and adjustment its execution, and the endorsement (or rejection) of its informational deliverances. 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’.

11 Something along these lines is also suggested by Palermos (2015). He comments that “It is the subject’s organismic [i.e., biological] cognitive faculties that are first and foremost responsible for the recruitment, sustaining and monitoring of the extended belief-forming process...in virtue of which the truth...is eventually arrived at” (p. 2963).
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., the chances that it will 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 it’s 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 the extended process that 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, in press).

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 forms of technological extension? The answer to this question is relevant to the issue of whether a weakened form of virtue reliabilism (viz., COGAWEAK) establishes something of a snug fit with active externalism (Pritchard, 2010). If the metacognitive processes are not, typically, extended, then there 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 favour of COGAWEAK.

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 COGAWEAK 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 COGAWEAK 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 the material realization of such processes, to my mind, is along the lines of the predictive processing model that is described by Clark (2013, 2016). In fact, in a recent paper, Clark (in press) makes an explicit link between neurally-grounded predictive processing mechanisms and metacognitive capabilities. The kinds of processes in question are thus ones that, for the most part, are 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, 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 (in press) usefully notes “...precision estimation itself can be partially outsourced, scaffolded, and amplified using bio-external tools and resources”.

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.

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.

**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 in relation to philosophical debates concerning 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 regarding the nature of knowledge or the potentially-extended character of human cognition – there is, for example, no reason to doubt claims regarding the possibility of Web-extended cognizers. Neither 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 might undermine 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 extended cognizers. In order to make sense of this claim, let us focus our

---

12 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.
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 the 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 & Record, 2013; Simpson, 2012). Consider, for example, the case of personalized search results. 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 seen as leading to ‘filter bubbles’ (Pariser, 2011). The epistemological significance of such filter bubbles, according to Miller and Record (2013), is to limit the justificatory status of a user’s beliefs and threaten their overall epistemic standing. Inasmuch as such claims are true, it seems that we confront a situation where 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 a make a technology apt for cognitive incorporation are not always 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 & 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, 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, p. 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 judgements regarding the truth status of processed information (Reber & Schwarz, 1999; Reber & 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,
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 accept that 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 serves as a source of concern regarding 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 becomes apparent. Recall that in discussing linked data, I sought to highlight the value of linked data formats in terms of 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; however, 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 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 judgements 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 judgements, 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 knower seems to be undermined.

In closing this section, I want to highlight another 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) at an unprecedented rate. It is also a concern that is born out of a specific focus on virtue reliabilistic theories in contemporary epistemology. In order to better understand the 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 technologies, I suggested, seem to satisfy many of the conditions that guide our intuitions concerning when we confront a genuine case of extended cognition. 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 terms of establishing positive epistemic status. Abilities, however, are the sorts of things that typically take time to acquire\textsuperscript{13}. 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. Given that rapid rates of technology 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 may come in a variety of forms, but undoubtedly the most epistemically-pernicious kind of change occurs when the operation of a technology is changed in a way that is largely invisible to the end user. Software upgrades that affect the behind-the-scenes behaviour 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\textsuperscript{14}, 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.

**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 focusing on emerging digital technologies is that they

\textsuperscript{13} This is an assumption that seems to follow from our notion of what abilities are. One often encounters claims of this sort in the epistemological literature. For example, Palermos (2013) 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”.

\textsuperscript{14} Clark (in press) also alludes to a ‘hidden vulnerability’ in accounting for our reluctance to credit agents with knowledge.
highlight a range of issues that might not otherwise come to light if we limit ourselves to the rather mundane and technologically-low grade examples favoured by the philosophical community.

My analysis of emerging digital technologies suggests that the general thrust of technology design is largely consistent with the kind of criteria that have been used to guide our intuitions as to when we confront a genuine case of cognitive extension. In fact, if anything, the features of emerging digital technologies seem to make them better candidates for cognitive incorporation, especially when compared to the kind of artefacts that typically animate philosophical discussions in this area.

When it comes to virtue-theoretic conceptions of knowledge we saw that notions of virtue reliabilism are largely compatible with the possibility of extended cognitive systems. 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 behaviour. 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 delivered information can exert over behavioural output. Given that these processes are typically realized by neural mechanisms, there does not seem 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 in 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 & 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 they 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 better able 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.

**Acknowledgements**

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.

References


Smart, P. R., & Shadbolt, N. R. (2014) Social Machines. In M. Khosrow-Pour (Ed.), *Encyclopedia of Information Science and Technology*. IGI Global, Hershey, Pennsylvania, USA.


