Forthcoming in *Extended Epistemology,* Carter, A. J., Clark, A., Kallestrup, J. Palermos, S. O., Pritchard, D. (eds.), Oxford University Press.

NEW HUMANS?

ETHICS, TRUST AND THE EXTENDED MIND

J. Adam Carter, Andy Clark & S. Orestis Palermos[1]

University of Edinburgh

Abstract. The possibility of extended cognition invites the possibility extended knowledge. We examine what is minimally required for such forms of technologically extended (and distributed) knowledge to arise and whether existing and future technologies can allow for such forms of epistemic extension. Answering in the positive, we explore some of the ensuing transformations in the ethical obligations and personal rights of the resulting 'new humans.'

1. What is Extended Cognition?

According to one traditional conception of mind and world, one that owes primarily to Descartes, the mind and its cognitions are regarded as entirely insulated from the external world which we aspire to know. We can think of this picture as one of an 'inner' and 'outer' world, where the two simply do not mix.

In the second half of the 20th century, Hilary Putnam (1975), Tyler Burge (1986) and Saul Kripke (1980) highlighted one important reason why this segregated picture of cognition and world can't be right. Under the banner of *content externalism* these philosophers established a point now widely taken for granted, which is that what one counts as thinking *about*—that is, the content of the thought you are currently having—

[1] Author names appear in alphabetical order.

is at least *partly* a matter of your physical or social environment, and thus can't be entirely a matter of how things stand internal to you.

While allowing the world to play a role in *individuating* mental content represents one form of resistance to the mind/world boundary, a much more provocative form of resistance emerges from a recent strand of thinking in the philosophy of mind and cognitive science. This strand of thinking, which has become known under the banner of the 'extended mind' depicts certain forms of human cognizing as inhering in complex tangles of feedback, feed-forward and feed-around loops that promiscuously criss-cross the boundaries of brain, body and world (see Clark and Chalmers 1998; Clark 2008; and essays in Menary 2010).

To make this idea more concrete, suppose you—perhaps, in order to guard against the early onset of Alzheimer's—begin relying on a user-friendly note-taking app on your smartphone for information encoding, storage and retrieval. When you learn something new, you record it in the 'memory app'; when you need old information, you automatically and unreflectively access the app. We can make this story subtler and more powerful by adding further capabilities to the app. Every time you look up one entry, the app now suggests similar or related information you might want to take a look at. The entries that have not been used for a long time fade out from the suggested list, and the ones that are most commonly invoked appear on the top. Perhaps the app can even track your current location and automatically project previous entries related to that location or the type of event you are attending. The app also automatically creates connections between the various entries: connections that, subject to the frequency they are being followed, get stronger or weaker. In fact, all this may soon be possible given the advent of the Semantic Web (Berners-Lee et al., 2001). Over time, relying on the super-app becomes automatic and second nature to you. The app has in effect started playing the functional role of information encoding, storage and

automatic when-needed retrieval. This is strongly reminiscent of at least some aspects of the functional role ordinarily played by *biological* memory.

Given such a well-integrated functional role, why insist that all your 'real' memory is in your head? There comes a point, Clark and Chalmers argue, where simply *insisting* that all processes of real human memory play out entirely within skin and skull looks like an unprincipled kind of '*bioprejudice*'. Proponents of *extended cognition* think we should shun bioprejudice for a more egalitarian approach to thinking about our mental lives. To drive this home, Clark and Chalmers suggested a so-called 'parity principle'—viz., a rule of thumb that states that:

> *Parity principle*: If, as we confront some task, a part of the world functions as a process which, were it to go on in the head, we would have no hesitation in accepting as part of the cognitive process, then that part of the world is part of the cognitive process. Clark and Chalmers (1998, 8).

Clark and Chalmers' claim is thus that when we think about 'the mind' we are often over-impressed by the ancient boundaries of skin and skull. Just as our physical capacities can be repaired, augmented, and transformed by new non-biological tools and technologies, so (the 'extended mind' story claims) can our mental capacities.

2. A Puzzle for Extended Knowers

In one way, all this can seem blindingly obvious. We use calculators, notebooks and iPads, and clearly we can do more search, reasoning, recall and calculation as a result. But in some cases we have the intuition that it is YOU, the agent, that has been enhanced, augmented, or extended, while in other cases it seems more like YOU, the agent, are accessing or deploying some additional, external, resource. Call these the 'extended' versus 'merely tool-like' cases respectively. Classic cyborgs, like Terminator and Robocop, fall clearly on the extended side of this divide. For the classic cyborg, the implanted technology helps make them the specific agent that they are. Old-fashioned

VCR remote controllers, on the other hand, leave the VCR clearly on the other side of the agent-world divide. These are intuitively best seen as (rather hard to use) tools rather than true user-augmentations. But the moral of the extended mind arguments (see also Clark (2003)) is that the difference between 'merely tool-like' and agent-extending technologies does not require mind-extending stuff to be wired directly to the brain, or even to be permanently attached to, or implanted in, the body. Instead, the claim is that if some additional non-biological resource is generally available when needed, fluently (and pretty much automatically) deployed, and mostly unreflectively trusted in what it delivers, then it already shares much of the bedrock functional profile of your native mental resources (see Clark & Chalmers 1998).

Those demands (availability, fluency, and trust) are meant to evoke a familiar profile. After all, we don't (mostly) stop and think "Hmm, that might be stored in my bio-memory…" then retrieve the info and carefully check it. Instead, bio-memory is (mostly) there when needed, and we just rely on it as we engage in a task. Truly mind-extending/cyborg technologies, Clark and Chalmers argue, need to be invoked and relied upon just as easily and unreflectively as we invoke and rely upon bio-memory, bio-reasoning, and bio-sensing. Similarly, a tennis racquet, in fluent use, gets folded right into the flow of sensing and acting, so that when we encounter the 'tennis world' we do so with the racquet playing a role more like that of a temporary body-part than that of an encountered tool or object.

But all this has an interesting consequence. It means that in normal operation, mind-extending tools should seek to by-pass the epistemic gatekeepers of deliberate, conscious, slow, careful, agentive attention. The best new-you-bits need to join the 'cognitive party' without being constantly stopped at the sensory gates and asked to show their invitations and IDs! This poses a puzzle. For it means that the best cognitive extensions will be, *prima facie*, among the worst cases of augmentations apt to deliver genuine knowledge This is a significant threat that can emerge from a number of

approaches to knowledge that put forward some form of 'awareness condition' on justification.

Take for example the standard epistemic internalist approach,[2] according to which one should always be able, at least in principle, to access the reasons that justify one's beliefs, *by reflection alone*.[3] It is difficult to see how such an approach to knowledge could line up with the possibility of knowledge produced on the basis of an extended cognitive process. If knowledge requires the ability to reflectively (i.e., consciously) access the reasons that support one's beliefs, then it seems highly unlikely that one could count as knowing a piece of information, which is unreflectively trusted and which has been automatically retrieved.[4] Or consider a strong version of the externalist approach of virtue reliabilism (Pritchard 2010), according to which the *agent* must be *primarily* creditable with the cognitive success of believing the truth. What that involves is not merely that the new-found capacity be reliable, simply as a matter of fact. The new capacity needs to be, in some intuitive but surprisingly elusive sense, *her own*. It needs to be primarily creditable to her (cognitive) agency. One way in which this requirement is standardly met is by insisting that the agent be aware of the source of the reliability of her new ability. Once again, however, the problem is that such a requirement stands in immediate tension with the suggestion, in the literature on the 'extended mind', that true cognitive extensions function automatically and (where possible) below the level of active agential scrutiny. Clark (2015) presents this as a dilemma, which we shall label the Epistemic Hygiene Dilemma:

[2] For an overview of this position and its key contrast points with epistemic externalism, see Pappas (2014).

[3] For classical defenses of this view see Chisholm (1977) and BonJour (1985, Chap. 2). See also Steup (1999), Pryor (2001, p. 3), BonJour (2002), Pappas (2014), and Poston (2008).

[4] For a recent argument in favour of the compatibility of epistemic externalism and the extended mind, see Carter and Palermos (2016). Cf., Smithies (2017, this volume) for a different perspective.

> *Epistemic Hygiene Dilemma.* Either the agent consciously encounters some new resource as an ongoing object for various forms of epistemically hygienic practice (such as understanding why it is a reliable source of information) or not. If she does, this makes the resource look, at that moment, more like external equipment (it may then be a source of knowledge while failing to be part of HER). If she doesn't, it looks unable (from the perspective of theories of knowledge that subscribe to some form of 'awareness condition') to act as a source of knowledge.

3. Assessing the Damage

How damaging is this to the case for extended knowing? A tempting move is to dismiss the worry as one that would equally well apply to the very best biological cases—a kind of reductio of the 'awareness condition'. After all, I may surely rely on operations provided by a pocket calculator as a source of mathematical knowledge even without knowing how the calculator works. And were such a device to become fluently assimilated into my daily routines, deployed unreflectively and automatically when the task demands it, that does not obviously change that situation in any material way. Such cases (seen from the perspective of a suitably weakened form of virtue epistemology[5]) may instead involve a kind of socially embedded epistemic vigilance. Some folk know how these things work, after all.

Unfortunately, however, many of the most powerful near-future tools and technologies will rely on stuff that *nobody* fully understands. A timely example concerns the use of so-called 'deep architectures'—viz., multi-level artificial neural networks used to discover (via 'deep learning') patterns in large data-sets such as images and texts. The power and prevalence of these deep architectures masks a major problem—the problem of *knowledge-opacity*. Such architectures learn to do wonderful things, but they do not (without further coaxing) reveal just what they are 'thinking about' when they do them.

[5] See, for example, Pritchard (2010).

A stark example of the limits of our understanding was the recent discovery that trained-up networks systematically admit of so-called 'adversarial exemplars'. Thus suppose a net has learnt to classify images as cars or as not-cars. In a piece with the understated title of 'Intriguing properties of neural networks', Christian Szegedy et al. (2013) showed how to generate (reliably and systematically) very subtly deformed 'adversarial' versions of images that the net would easily classify correctly. These adversarial versions involve small pixel-level perturbations that are invisible to the human eye, so the adversarial exemplar looks, to us, like an identical image. Yet the network gets them totally wrong. Just examining the networks normal behavior would have led us to conclude that it knew what cars looked like in general—and no-one would predict that a visually indistinguishable version of a photo easily classified by humans as a car would be so challenging as to defeat the network.

The good news is that networks can then be trained on adversarial examples to help combat those blind spots. But the larger moral is that deep nets learn ways of solving problems that are opaque to their human developers. This has been a long-standing problem for multi-level networks. But as the tide of deep learning advances, it is one we can no longer afford to ignore. Deep learning and the patterns it extracts will soon permeate every aspect of our daily lives, from online search, to online recommendation systems, to bank-loan applications, healthcare and dating. Appeals to socially distributed forms of epistemic virtue will not save near-future augmentations that exploit these kinds of algorithms from the dilemma of epistemic hygiene.

Another attempt to limit the damage immediately suggests itself. For plausibly, we do not know our own biological capacities any better than we know the new deep learning capacities just described. But our senses can deliver knowledge despite this. Yet with bio-sensing (sight, hearing, etc.) we at least have some good background reasons for trust. For one thing, we have the long history of evolutionary 'test'. In addition, we have our own extensive 'user-experience', honed during slow childhood

acclimatization. Most recently, we benefit from hundreds of years of careful scientific probing and research (revealing e.g., cognitive biases, and our proneness to certain illusions). But here too, the new technologies score rather badly. For the ever-changing suite of apps that might help constitute a cognitively extended 'new human' we have maybe a year or two of software development, a few beta-testers, and the (quasi-'evolutionary') credentials of being made by Google, Microsoft, or Apple. This should be worrying. Moreover, and fully in line with the basic tenets of extended cognition, the best personal upgrades will be readily poised for fast fluent assimilation into our daily routines. They will be upgrades designed in ways that increasingly preclude reflective attention and interrogation on the part of the newly enhanced agent. No-one wants to have to learn to use a new app. They just want it to work.

So far, then, the dilemma of epistemic hygiene is not resolved. The best cyborg/ extended-mind technologies will not lend themselves to 'proper epistemic care and vigilance' on the part of the agent. They are designed for rapid assimilation, not to present themselves as objects of concern. And the factors that may mitigate this worry in other cases do not seem to get an equal foothold with respect to ill-understood human-built technologies. How, then, are we to become epistemically responsible cyborgs?

4. A minimalist approach to epistemic hygiene

To answer this question, we can focus on a recent virtue reliabilist proposal that construes epistemic responsibility and agency in a rather weak fashion.[6] We previously noted that a number of epistemological proposals regard epistemic justification as involving the ability to provide explicit positive reasons in support of our beliefs or in support of the reliability of our belief-forming processes (i.e., what we previously called

[6] For some notable defences of virtue reliabilism in epistemology, see, for example, Sosa 1988; 1993; 2007; 2015; Greco 1999; 2010.

the 'awareness condition'). While this is an intuitive way to think about justification, the problem is that there are several belief-forming processes, such as biological vision and bio-memory, which are supposed to be knowledge-conducive, even though most epistemic agents have no idea how they work or why they are reliable. Accordingly, when we acquire knowledge on their basis, it seems incorrect to require explicit positive reasons in their support. To solve this long standing problem, a few epistemologists have recently suggested that we should give up the aforementioned strong understanding of justification, according to which one should be able to provide explicit positive reasons in support of one's beliefs, and instead embrace one of several weaker alternative visions.[7]

According to one prominent weaker alternative, in order for one's true beliefs to qualify as knowledge, they must simply be the product of a belief-forming process that counts as a cognitive ability. This is known as the *ability intuition on knowledge* and its intuitive appeal comes from the fact that cognitive abilities do seem to be the sort of belief-forming processes that can generate knowledge, even if one has no explicit positive reasons to offer in their support.[8] No one needs to explain why their vision or hearing is reliable when they come to acquire knowledge on their basis, after all.

But if this is the way to approach knowledge and justification, there are at least two more central questions that we need to further ask: (1) When does a process count as a cognitive ability belonging to a certain agent and thereby as knowledge conducive, and—depending on how we answer (1)—(2) what is the sense in which one can be justified on the basis of one's cognitive abilities in a way that does not involve the possession of explicit reasons in their support?

---

[7] See, for example, Huemer (2007), Burge (1993), Pryor (2004), etc.

[8] The idea that knowledge must be grounded in cognitive abilities can be traced back to the writings of (Sosa 1988, 1993) and Plantinga (1993*a*, 1993*b*). For more recent approaches to this intuition, see Greco (1999; 2004; 2007; 2010) and Pritchard (2009, 2010, 2012).

In order to answer the first question, John Greco (e.g., 1999; 2010; 2012) insists that in order for a process to count as a cognitive ability, it must have been appropriately integrated into the agent's cognitive character, which he defines as the set of 'acquired skills of perception and acquired methods of inquiry including those involving highly specialized training or even advanced technology' (1999, 287). But what exactly is required in order for a process to be so integrated?

As far as common-sense intuitions are concerned, Greco (1999, 2010) has noted that the relevant belief-forming process must be neither strange nor fleeting[9] (i.e., it must be a normal, dispositional cognitive process). Nevertheless, in later work (2010), Greco's explanatory focus shifts away from such broad intuitions and towards a more mechanistic understanding of cognitive integration that centers around the cooperative interaction between the relevant process and the rest of the agent's cognitive character. Specifically, he writes: 'cognitive integration is a function of cooperation and interaction, or cooperative interaction with other aspects of the cognitive system' (2010, 152).

What might be the reason that Greco spells out 'cognitive integration' and 'cognitive character' in this alternative way? The answer, Palermos (2014) suggests, has to do with a *minimal notion of epistemic agency and responsibility*. Greco is after a notion of subjective justification/epistemic responsibility that is in line with epistemic externalism in that it denies that in order to be subjectively justified/epistemically responsible one needs to have access to or be aware of the reasons for which one's beliefs are reliable. Accentuating the integrated nature of one's cognitive character allows for just this: If one's belief-forming process cooperatively interacts with other aspects of one's cognitive system, then it can be *continuously* monitored in the background such that *if*

[9] This caveat was added in part as a way to defend against certain 'meta-incoherence' style cases, such as Alvin Plantinga's (1993a) 'brain lesion case', where an individual's brain lesion causes her to reliably believe she has a brain lesion, though the individual is unaware of the source of the belief.

there is something wrong with it, *then* the agent will be able to notice this and, if she or he is so inclined, respond appropriately. Otherwise—if the agent has no negative beliefs about his/her belief-forming process—he/she can be subjectively justified in employing the relevant process *by default*, even if he/she has absolutely no positive beliefs as to whether or why it might be reliable.

On this version of virtue reliabilism, provided that (1) the agent's belief-forming process is integrated to his/her cognitive character by being *continuously* monitored by (or interacting with) the rest of the cognitive system, and (2) the agent is conscientious (i.e., motivated to believe what is true) such that he/she would indeed be responsive *were he/she* to become aware that there is something wrong with his/her process, then the agent can be subjectively justified in holding the resulting beliefs merely by *lacking* any negative reasons against them.

4.1 How minimalist?

According to this weaker approach to justification, in order to be justified, one does not need to be aware of the source of the reliability of her beliefs. Instead, one need only be in a position to become aware (in the relevant counterfactual circumstances) that his/her beliefs are unreliably formed and then respond appropriately. But even this, it can be argued, is requiring too much by way of active agentive involvement.

According to the picture developed above, even though the agent does not need to be aware of any explicit *positive* reasons in support of his/her beliefs, she should at least be able to become *aware* of any telltale signs of unreliability of the relevant process or the resulting beliefs. But drawing on recent cognitive scientific work in the area of 'predictive processing' and the 'Bayesian brain', Clark (2015) argues that an agent may still be epistemically responsive even if he/she *never* becomes consciously aware of any signs of unreliability. For example, driving on a foggy day, our sensory processing systems may automatically decrease the weight assigned to visual cues, while

increasing that assigned to prior knowledge of the road, auditory signals, etc. Conversely, on a clear and sunny day, our predictive brain increases the weight assigned to the incoming visual information. Crucially, in both cases, the 'Bayesian' predictive brain does so automatically, with no conscious decision required on the part of the agent. In the class of models Clark considers, these adjustments emerge from a process that continuously estimates the reliability or unreliability, in context, of specific beliefs and sources of information. Crucially, these estimations are part-and-parcel of the process of using incoming sensory information to perceive and act in the world, and can take place without any form of conscious awareness of the altered balances of power (e.g., between visual and auditory information, or between sensory evidence and top-down prediction) involved.

This is not to claim that the ability to become *aware* of these patterns of reliability and unreliability plays no role in the process of acquiring knowledge. Instead, the claim is that these kinds of sub-personal goings-on already satisfy minimal requirements on the epistemically sound (and hence agent-creditable) deployment of knowledge and evidence.[10]

But what about cases where we seem to lack sufficient information to make a choice, or perform an action, at all? Sub-personal forms of epistemic 'balancing' might be good for automatically adjusting the reliability of our belief-forming processes so that they can pass the most minimal threshold for knowledge.[11] Nevertheless, an agent that is capable of conscious awareness can go further. She can actively decide to pay greater attention to a tricky epistemic task, or decide that her grip on the situation is just too frail to be allowed to select any positive action at all—for example, parking her car

[10] How can something be credited to the agent if the agent need not 'do it' (even counterfactually)? The suggestion is that these capacities of sub-personal self-correction partially define the agent as the agent she is. I can be credited with them in the same way as I can be credited with having a certain bedrock temperament, or tendency to embrace or avoid risk.

[11] Compare with Sosa's (2015, Ch. 3, e.g., 76) notion of 'subcredal animal knowledge'.

by the roadside until the fog clears. Notably, conscious, *thoughtful reflection* on the part of the agent is not obviously necessary to make such choices;[12] For example, non-human animals may automatically estimate that they have insufficient information to proceed with a pounce-attack. Nonetheless, being capable of being consciously aware of the reliability (or at least unreliability) of one's cognitive capacities is not the same and it is not as epistemically demanding as being capable of human-level, *reflectively thoughtful* deliberations.

Our human capacities for reflection are admittedly causally potent, and enable us to exert caution in a much wider range of circumstances, drawing on shared knowledge of past outcomes, and complex imaginative simulations of possible future unfoldings. An additional role for conscious thoughtful awareness, Clark (2015) suggests, emerges not during the process of forming our beliefs, but during the process of *building* our belief-forming processes: 'consciousness is a necessary condition of a form of self-alienation—a form of self-alienation that opens up the space for all these more deliberate forms of epistemic engineering' (2015, 3773). It allows us to treat our thought processes as objects of other ongoing thought processes with the attendant potential to refine and develop them in ways that would perhaps be unavailable in the absence of conscious awareness. 'There is a close analogue in sports skills, where only reflective agents can ask what aspect of their golf swing (for example) was at fault when the ball lands in the rough', (*Ibid.*, 3773).

A minimalist approach to epistemic hygiene could therefore suggest that conscious agentive control of our beliefs may come synchronically (by being in a position to become aware of any potential faults with our beliefs as they are being formed), diachronically (by having been previously involved in the shaping of the

[12] Of course, sometimes due to practical considerations, we are forced to act even on the basis of impoverished evidence. Even in such cases, epistemic hygiene can be better or worse. For discussion on such cases of 'virtuous guessing' see Carter, Jarvis and Rubin (Forthcoming) and Carter (Forthcoming).

belief-forming processes that automatically generate our beliefs in the here-and-now) or even both. Admittedly, intuition can cut both ways. But all these approaches to minimalist epistemic hygiene share the following commitment: In order for the agent to count as epistemically responsible, what is minimally required is *responsiveness* to the contextual reliability or unreliability of the relevant belief-forming processes. So long as the agent can respond to the potential unreliability of her belief forming processes (independent of whether such responsiveness involves conscious awareness at the time of belief formation), the agent can come to acquire knowledge. This will be so even if he/she lacks *any* positive reasons to offer in support of the reliability of his/her belief forming process. On both approaches, this responsiveness arises out of the integrated nature of the agent's cognitive system, and as such it is a form of responsibility that can be properly ascribed to *her*—it is *that agent's* responsibility as it arises out of the interconnectedness of that agent's cognitive system as a whole.

4.2 Subliminal Belief-Forming Processes, Extended

It is therefore possible to understand epistemic agency and responsibility in a way that allows one to acquire knowledge on the basis of belief-forming processes such as vision, memory and hearing, even if one could have never become aware of any non-circular, explicit reasons to offer in their support. With regards to the dilemma of epistemic hygiene, this approach to epistemic agency and responsibility seems to allow for the acquisition of knowledge on the basis of a belief-forming process that is extended, even if the relevant extended process is automatically deployed, unreflectively trusted and, on the whole, operates *subliminally* in a way that parallels the functioning of most of our biological cognitive resources. In fact, there are some interesting parallels between the way epistemology approaches the idea of cognitive integration and the literature on extended cognition.

Arguments for extended cognition do not always rest on the commonsensical intuitions that we considered in §1—i.e., the fluency, automaticity and unreflective reliance of the agent on the external resources. Instead, a number of theorists have recently suggested focusing on the details of Dynamical Systems Theory (DST) (Clark and Chalmers 1998; Chemero 2009; Froese et al. 2013; Palermos 2014).

Dynamical Systems Theory (DST) is a successful and widely used branch of theoretical mathematics that deals with the study of systems that change over time. According to DST, whenever some process is performed by two or more components that interact reciprocally with each other, so that (for example) component one causes changes in component two, while component two causes changes in component one, the resulting behavior can be fruitfully described as arising within a single (coupled) system. Many instances of current or near-future cognitive technologies will display just such a profile. This provides further reason to analyse such cases by identifying an extended cognitive system that comprises the biological agent along with all such properly coupled elements. Such couplings may also characterize a range of other cases, including the use of Tactile Visual Sensory Substitutions Systems),[13] and cases in which two or more agents collaboratively perform a cognitive task, such as performing a scientific experiment. This latter possibility—known as *distributed cognition*[14]—may sound even more radical than the possibility of cognitive extension, but the only difference is that, this time, the extended cognitive system putatively extends to include not only artifacts but other biological agents as well.

Given this understanding of extended and distributed cognition as arising out of the synergetic operation of two or more coupled components and the fact that a similar ongoing interdependence is required between the components of a cognitive system to

[13] See Bach-y-Rita and Kercel (2003) for a recent review on TVSS.

[14] Hutchins 1995, Sutton, 2008; Sutton et. al, 2008; Theiner et al., 2010; Theiner and Goldstone, 2010; Palermos Forthcoming*a*.

count as knowledge-conducive, it seems that we have an approach to knowledge and justification that allows for the acquisition of knowledge on the basis of not only subliminal biological cognitive processes, but also subliminal processes that are technologically extended or even distributed (we return to this latter possibility in section 5.3). On this view, an extended belief-forming process may belong to an agent *S* who can take epistemic responsibility for its outcomes, even if its operation is entirely automatic and unreflective.

As noted in the previous section, this is (i) either because the agent can at least become counterfactually aware of any signs of unreliability and thereby choose to consciously respond appropriately should there be anything wrong (even though, when everything goes well, the agent does not need to exercise any conscious control at the time of forming her beliefs), (ii) because the agent has, in the past, been consciously involved in the epistemic engineering of the belief-forming processes that automatically generate her beliefs (even though at the moment of forming her beliefs she needs to have no factual or counterfactual awareness of their reliability), or possibly (iii) because of both (i) and (ii). The common denominator behind all these minimalist approaches to epistemic responsibility, ownership and overall hygiene is that they all account for the indispensable epistemic requirement that the agent's belief-forming processes be responsive to their own contextual reliability, and in all cases this responsiveness arises out of the integrated nature of the agent's cognitive system as a whole. Granted, this is a minimalist way to understand ownership—and thereby responsibility—of a belief forming process, but it seems to be the same kind of responsibility that is minimally required in order to acquire knowledge on the basis of our bedrock biological resources.

For example, it is possible to use the above approach in order to explain how a subject might come to perceive the world on the basis of a Tactile Visual Substitution System (TVSS), while also holding fast to the idea that knowledge is belief that is true in virtue of *cognitive ability* (i.e. the ability intuition on knowledge). Briefly, a tactile visual

substitution system is a mini video camera attached on a pair of glasses, which converts the visual input into tactile stimulation under the agent's tongue or her forehead. By moving around and on the basis of the associated sensorimotor contingencies,[15] blind patients quickly start perceiving shapes and objects and orient themselves in space. Occasionally, they also offer reports of feeling as if they are *seeing* the objects, indicating that they are enjoying phenomenal qualities that are close to those of the original sense modality that is being substituted. Seeing through a TVSS qualifies, in the light of DST, as a case of cognitive extension, because it is a dynamical process that involves ongoing reciprocal interactions between the agent and the artifact.[16] By moving around, the agent affects the input of the mini-video camera, which continuously affects the tactile stimulation she will receive on her tongue or forehead by the TVSS, which then continuously affects how she will move around and so on. Eventually, as the process

[15] For a full account of how sensorimotor knowledge is constitutive of perception see (Noë 2004). "The basic claim of the enactive approach is that the perceiver's ability to perceive is constituted (in part) by sensorimotor knowledge (i.e. by practical grasp of the way sensory stimulation varies as the perceiver moves)". (Noë 2004, 12). "What the perception is, however, is not a process in the brain, but a kind of skilful activity on the part of the animal as a whole". (Noë 2004, 2). "Perception is not something that happens to us or in us, it is something we do". (Noë 2004, 1). Sensorimotor dependencies are relations between movements or change and sensory stimulation. It is the practical knowledge of loops relating external objects and their properties with recurring patterns of change in sensory stimulation. These patterns of change may be caused by the moving subject, the moving object, the ambient environment (changes in illumination) and so on.

[16] One possible worry is that ongoing mutual interactions may not be sufficient for cognitive extension. Instead, general availability of the external resource might also be crucial in the same way that our biological cognitive resources are readily available almost anywhere we go. Indicatively, Wilson and Clark (2005) consider the possibility of TECS—transient extended cognitive systems—but they treat them in a rather sceptical manner, "for we properly expect our individual agents to be mobile, more-or-less reliable, bundles of stored knowledge and computational, emotional, and inferential capacities, and so we need to be persuaded that the new capacities enabled by the addition of the notebook are likewise sufficiently robust and enduring as to contribute to the persisting cognitive profile we identify as [the agent]." While it is true that general availability seems to play an important role when we intuitively judge whether a specific case of tool-employment should qualify as a case of cognitive extension it is not entirely clear that it is a necessary condition on cognitive extension or that it is a condition that should be thought in relation to the relevant organismic agent. Using pen and paper to solve a mathematical problem probably qualifies as a case of cognitive extension even if most of us do not normally carry a pen and paper around everywhere we go. Similarly, our biological cognitive resources might be generally available to us but not all the time; for example our visual perceptual capacities are unavailable whenever lighting conditions are low. It might therefore be possible that general availability is a background condition pertaining to the environment that the agent is normally embedded in (for example that normally it is easy to get hold of pen and paper or that lighting conditions are most of the time sufficient for visual perception) rather than a condition on the agent-artifact system itself.

unfolds, the coupled system of *the agent and her TVSS* is able to identify—that is, see—shapes and objects in space.

Seen from the perspective of virtue reliabilism, the belief-forming process in virtue of which the subject believes the truth with regards to the space surrounding her might indeed be for the most part external to her organismic cognitive agency, but it still counts as one of her cognitive abilities, as it has been appropriately integrated into her cognitive character. It is the ongoing interplay between the agent's organismic cognitive faculties and the working of the external component that is responsible for the recruitment, sustaining, and monitoring of the extended belief-forming process (i.e., quasi-visual perception), in virtue of which the truth with respect to shapes and objects in space is eventually arrived at in a reliable fashion.

5. The extended knower: ethical implications

With the initial dilemma of epistemic hygiene finally resolved, the possibility of *bona fide* extended knowers has now been established. In what follows, the aim will be to look beyond the epistemology of 'new humans', by examining in some detail the kind of 'extended ethics' that would characterize the extended knower (in the sense described in §§1-4). Consider that, at a high level of abstraction, what we should do often depends both on what we know and on what we're capable of finding out. *Extended knowing*, in contrast with traditional intracranial knowing, represents a shift in both the breadth of the information base we're capable of commanding as well as the comparative ease by which new knowledge can be acquired, e.g., through responsible use of cyborg/extended-mind technologies. It should not be surprising, then, if certain aspects of our ethical and legal thinking that depend upon what we know and what we know how to know should require some updating.

In this section, we will suggest three concrete ways in which extended knowing gives rise to new ethical and legal challenges: specifically, these concern (i) expertise and ethical obligations; (ii) privacy; and (iii) responsibility.[17] We consider each in turn.

5.1 Expertise and Ethical obligation

Here is a very intuitive principle: *ceteris paribus*, if you do *not* know how to φ, you are *not* obligated to φ. You can't be obligated to prove the Riemann Hypothesis if you failed high school math. The converse, of course, is false: *knowing how to* φ is not sufficient for generating an obligation to φ. You might know how to build a bomb; that does not mean you should. However, the possession of the right kind of knowledge can, as Vanessa Carbonell (2013) puts it, 'trigger' ethical obligations, under certain circumstances where other conditions for ethical obligation are already satisfied. Call this the 'triggering principle':

> *Triggering Principle*: knowledge "triggers" obligation when it provides one of the necessary and jointly sufficient conditions for obligation—specifically, when it is either the last (temporally) of the conditions to obtain, or when it is the condition that sets a given agent apart from some comparison class (2013, 246).

Especially interesting for our purposes will be the latter kinds of cases, where knowledge possession triggers an ethical obligation by setting the agent apart from a comparison class, namely, individuals who *lack* the relevant knowledge. Carbonell offers a helpful illustrative case, which we can use as our reference point:

> …[S]uppose a man collapses on the railway platform and is dying while waiting for the paramedics. As the sole bystander I would be obligated to save his life,

[17] A further important implication of extended epistemology for ethics and law concerns the legal definition of assault. Of particular relevance is that the intentional compromising of an individual's faculties or powers is ordinarily conceived of as a kind of personal harm distinct from (say) merely damaging that individual's property. However, if we leave the intracranial picture of the mind behind, there is room to consider cases where an individual's faculties are intentionally compromised by the causing of harm or damage to the extra-organismic material realisers of such faculties. For discussion on this point, see Carter & Palermos (Forthcoming).

> but I do not know how. (Fortunately, the paramedics arrive just in time.) Coincidentally, a CPR course is offered at my workplace that day, and I take it. On my return commute, shockingly, another man collapses on the railway platform. No one else on the platform has the relevant knowledge, but now I do. A knowledge-based obligation has been triggered (2013, 247).

In this case, on the return trip home (after taking the CPR class), what triggered the obligation was simply *knowing* the steps to perform CPR to save a life. In the above example, the knowledge in question is presumably stored in bio-memory, which the individual in question can easily recall (having just learnt the information).

Contrast now the above case, where the triggering principle kicks in, with an 'extended variation' on the case. Let's tweak the details of Carbonell's scenario so that the subject in question—let's call her C.P.—never took the CPR class that day. However, suppose the she instead simply purchased and downloaded the 'CPR Tempo[18]' app for her iPhone, which offers both audio as well as visual cues that aid the timing of chest compressions during the process of cardiopulmonary resuscitation (CPR). According to the American Heart Association, approximately 100 to 120 compressions per minute are required in order to administer CPR effectively.[19] Let's suppose that C.P. appreciates that, but given that she never took the class, she would be hopeless at administering CPR without the audio and visual timing cues of the CPR Tempo app.

Let's continue our variation on Carbonell's case; in particular, let's hold fixed that no one on the train platform (other than C.P., whose phone now has the CPR Tempo app installed) has any clue about CPR performance. With reference to the triggering principle, it looks as though C.P., *despite* lacking any relevant information about CPR in bio-memory, nonetheless now has a knowledge-based obligation she would otherwise lack. Specifically, this is an obligation that becomes triggered by her

[18] https://itunes.apple.com/gb/app/cpr-tempo/id525695057?mt=8

[19] These are the updated guidelines since 2015, prior to which the suggested number was 100. http://news.heart.org/%EF%BB%BFnew-resuscitation-guidelines-update-cpr-chest-pushes/

having integrated (and let's suppose, she practiced using it several times after downloading) the CPR Tempo app into an iPhone she's already very fluent with.

If the foregoing diagnosis is right, then it looks very much as though extended knowers will, on the whole (and abstracting now from the particular CPR case), be more likely to satisfy epistemic triggering conditions on various kinds of obligations—obligations they would fail to have in the absence of cyborg/ extended-mind technologies. This is, we think, the right conclusion to draw, and it is an important ethical consequence of extended knowing.

Consider, however, one kind of anticipated objection, according to which our diagnosis of the case of C.P. implausibly overgeneralises: 'C.P., in this redescribed case, merely *knows how to find out* how to do CPR properly. But, there is a sense in which everyone else on the platform also knows *how to find out* how to do CPR (i.e., by asking an expert, looking it up, downloading the CPR Tempo app themselves, etc.). Thus, the fact that C.P. knows how to find out how to do CPR does not *set her apart* from the comparison class that is the other individuals on the platform. And therefore, the obligation is not triggered.'

In reply, consider that the kind of knowledge that is plausibly relevant to the triggering principle is actionable knowledge in the context of the obligation, that is, knowledge that can be promptly put to use. If someone—call her Mnemony—possessed all the CPR-relevant information in bio-memory, but before putting it to use had to first recite 35 detailed and extremely slow mnemonic devices (all stored in bio-memory), her knowledge would not count as actionable knowledge. And this is the case even if we insist that Mnemony does know how to perform CPR, but has to slowly and painstakingly work through her mnemonic devices first. In our variation on Carbonelle's case, C.P. plausibly has an obligation triggered that Mnemony does not, despite Mnemony's possessing the relevant knowledge.

Furher complexities could ensue if others present had even faster and more powerful devices and greater expertise at following newly downloaded instructions. In such an instance, it could be that one of the 'merely potential' extended knowers—the ones who could, if they wished, download the app and apply the lessons rapidly—might have the strongest obligation to step in! [20]

Given increased cognitive offloading and seamless integration with iPhones, Satnavs, smartwatches and the like, 'lack of knowledge'—as an exculpatory factor in cases where other conditions for ethical obligations are met, will plausibly become less common. In situations where an individual is in need and several bystanders have smartphones, the condition that sets a given agent apart from some comparison class (and thus triggers the knowledge-based obligation) might well come down to *fluency* of coupling, with considerations about who's stored what in their bio-memory fading to the background.

5.2 Privacy

Your thoughts are your own. As soon as someone tries to take them from you (imagine: your private thoughts are broadcast for millions to see) you are, in an important sense, less free and autonomous—or at the very least, more pragmatically constrained—than you were before. As Michael P. Lynch (2013) puts it:

[20] All these cases hide a complication, which is that they implicitly assume that all agents present can rapidly determine their relative epistemic status compared to all the other agents present, so that it is immediately clear to the 'fastest, best' agent that she has prime responsibility to act. This was relatively easy in the case of comparing bio-knowers ("is there a doctor or a para-medic on the train?") but our self-assessments of what we *could* come to know, and how easily, and with what reliability, are plausibly far less reliable. Perhaps developing this kind of meta-self-knowledge will be an important part of future ethical training.

> to be an autonomous person is to be capable of having privileged access … to information about your psychological profile—your hopes, dreams, beliefs and fears. A capacity for privacy is a necessary condition of autonomous personhood.

How safe are the contents of your mind? Recent revelations of the extent of governments' access to Big Data has made salient one glaring way that our private lives, as expressed online through revealed preferences and social media, are not as private as we thought.[21] Nowadays, this overarching threat to online privacy, in the name of public safety, is a heated point of political debate.[22]

However, there is a much subtler way in which extended knowers stand to have their privacy compromised by laws which presuppose that our private lives are—for legal purposes, at least—in our heads. Here it will be helpful to briefly review the 2014 landmark U.S. Supreme court ruling in *Riley v. California,* a case in which a California citizen's phone was searched, without a warrant, during the course of an arrest following a traffic stop.[23]

The lower courts which ruled in *Riley v. California* regarded the arresting officer's search to be lawful, because arresting police officers are allowed to perform, without a warrant, search of an arrested person's physical area, which is defined as the area within the person's physical control. This legal precedent in the U.S. is called 'Search Incident to Arrest' (SITA), also known as the Chimel rule.[24] Riley's mobile phone was within the area of his physical control at the time of arrest, and so was deemed within the bounds of what could be permissibly searched without a warrant. Incriminating

---

[21] Examples here include the FBI web surveillance system Carnivore which samples non-suspect internet communication, and the Echelon global satellite network. See DeCew (2013) for a philosophical overview[§4]. See also http://www.theguardian.com/technology/2014/jun/20/little-privacy-in-the-age-of-big-data.

[22] For discussion on the 'privacy/security' trade off in EU politics, see Mortera-Martinez (2015).

[23] For the U.S. Supreme Court opinion, see http://www.supremecourt.gov/opinions/13pdf/13-132_8l9c.pdf. For a discussion of this case in connection with the extended cognition debate, see Carter & Palermos (2016).

[24] https://supreme.justia.com/cases/federal/us/395/752/case.html.

evidence was subsequently found on the phone, and because the lower courts regarded the search to be lawful, this evidence was allowed to be used against Riley.

It should not be surprising that in the United States, the Chimel rule does *not* permit a warrantless search of an *individual*'s physical interior (i.e., an investigation of the contents of one's leg, arm, or brain). Nor does the Chimel rule permit the arresting officer to administer on-the-scene sodium thiopental (i.e., truth-serum), in an effort to somehow extract 'thoughts' without a warrant.

Problematically, though, for extended knowers, warrantless searching of a mobile phone (whether it is within the individual's physical area or not) is effectively a warrantless invasion of the mind, a point the Chimel rule glosses over by regarding physical objects in the arrested individual's area of control to be on an equal footing with respect to privacy.

Fortunately, laws are beginning to catch up (albeit slowly).[25] In 2014, the Supreme Court unanimously overturned the lower courts' rulings, and insisted that a mobile phone should be treated as relevantly *different* from other physical objects in the individual's areas of immediate control, and that as such should be exempt from what can be searched without a warrant. Chief Justice John Roberts, in his written majority opinion of the Court, unsurprisingly did not include any explicit commitment to the bounds of cognition in the course of his rationale. But by way of an analogy that featured in the majority opinion, the legal rationale did come close:

> '[…] modern cell phones . . . are now such a pervasive and insistent part of daily life that the proverbial visitor from Mars might conclude they were an important feature of human anatomy'.

The ruling in *Riley v. California* is a step in the right direction. However, many privacy laws remain tacitly wedded to the old intracranial picture of our mental lives

[25] For an extensive treatment of how the extended cognition hypothesis can impact on the law see Carter and Palermos (Forthcoming).

and the legal reasoning about privacy that falls out of it;[26] the need for further legal updates remains urgent. After all, as an expert panel at the Pew Research Centre predicts, in less than 10 years (i.e., by the year 2025), we will live in an

> 'environment where accessing the Internet will be effortless and most people will tap into it so easily it will flow through their lives "like electricity."' […] mobile, wearable, and embedded computing will be tied together in the Internet of Things, allowing people and their surroundings to tap into artificial intelligence-enhanced cloud-based information storage and sharing.[27]

5.3 Collective responsibility in cases of distributed cognition and Social Machines like Wikipedia. Distributed cognition, as noted before, takes place when two or more individuals engage in reciprocal interactions in order to solve some cognitive task (Palermos Forthcoming*a*). Perhaps the most well studied distributed cognitive systems are Transactive Memory Systems where two or more individuals collaboratively store, encode and retrieve information, thereby forming a collective memory system much in the same way that you and your memory app would (Wegner 1987). The idea of distributed cognition, however, has also started gaining traction within philosophy of science, and especially with reference to scientific research teams (Giere 2002, Palermos 2015), and it is particularly amenable (though so far largely underexplored) to Web science (Palermos Forthcoming*a*), and especially the case of Social Machines, such as Wikipedia—i.e., processes in which the people do the creative work and the machine does the administration (Berners-Lee et al 2002, 172) and which will enable to "do things we just couldn't do before" (Berners-Lee et al. 2002, 174).

[26] This is especially the case as concerns a British cyborg Neil Harbisson, whose 'Eyeborg' device, which allows him to perceive colour through sound saves, was damaged as if it were a camera by police officers.

[27] http://www.pewinternet.org/2014/03/11/digital-life-in-2025/

One of the most interesting questions concerning distributed cognitive systems and Social Machines is the issue of responsibility and attribution of credit. Given that in such cases the cognitive task is performed by the cognitive system as a whole, it would seem—contrary to standard practice—wrong, or unfair, to attribute the credit (or blame) to any individual alone, or even to a subset of them): The final product arises out of dense processes of interaction between the members of the relevant group, such that it may be impossible to isolate how the efforts of any given agent was involved.

Consider for example the case of English Wikipedia, which according to a recent study is almost as equally reliable as Encyclopedia Britannica (Giles 2005).[28] Contrary to Britannica, however, which assigns the authorship of its entries to well-qualified contributors that work in isolation, the reliability of English Wikipedia is a collective variable that is the product of the 'power of the many eyes' (Noveck, 2007). In order to grow fast in as many directions as possible, Wikipedia has always operated on a policy of free editability, according to which anyone can edit without providing any epistemic credentials. Given, however, that anyone can edit, Wikipedia has drawn the caring attention of a huge volume of contributors such that anytime a mistake is spotted it gets almost immediately corrected. In this way, the probability that any given information currently posted online is reliable becomes high, and this ongoing reliability is a collective property in the sense that it does not arise from any individual's expertise or credentials. Instead, it emerges from the co-operative activity of all the individual members, much in the same way that the interconnectedness of our individual cognitive systems makes it possible to detect cognitive shortcomings as soon as they occur and respond appropriately.[29]

[28] At least with respect to a wide range of scientific topics. See also encyclopaedia Britannica's response (http://corporate.britannica.com/britannica_nature_response.pdf) and *Nature's* counter response (http://www.nature.com/press_releases/Britannica_response.pdf).

[29] For a detailed account of Wikipedia as a distributed epistemic agent see (Palermos Forthcoming*b*).

Indicatively, notice that even if a single individual intentionally posts a falsehood that somehow remains online long enough for others to (mis) use it , it would be subtly misleading to lay all the blame on the individual who posted the false information. It would be misleading because Wikipedia's reliability, as we saw, does not originate from the reliability of the source of the information but from the community's well-developed capacity to check and correct any information that is posted. So when unreliable information remains online long enough to cause serious issues, this is a failing of the Social Machine as a whole.

Wikipedia is an example that we all happen to be familiar with. But the advent of the Social Web and the increasing use of knowledge management systems suggests that these issues concerning collective responsibility will become increasingly important as we shape and are shaped by a seamlessly interconnected world.[30]

Conclusions

One of the characteristics of the most powerful emerging technologies is their almost invisible nature: Fast automated information-retrieval can deliver reliable outputs much in the same way that our onboard cognitive capacities of bio-memory and bio-perception do. To ensure that these devices behave like parts of us, rather than opaque outside forces, we need to ensure that they become cognitively integrated in the many ways we have outlined. Such integration may be brought about by an ongoing series of reciprocal interactions that result in the correct (automatic) estimation of their context-variable reliability. However it is achieved, cognitive integration ensures that these technologically augmented 'new humans' can count as knowing and epistemically responsible agents. Technological extensions of this integrated sort radically transform what we know. But such a reconceptualization of our knowledge capacities demands an

[30] For a detailed treatment of the general topic of the distribution of epistemic agency, see Palermos and Pritchard (Forthcoming).

accompanying reconceptualization of our human nature, our ethical obligations, and our personal rights and duties.

Acknowledgements

The paper was produced as part of the AHRC-funded 'Extended Knowledge' research project (AH/J011908/1), which was hosted at Edinburgh's *Eidyn* Research Centre.

References

Blitz, M.J. (2010). "Freedom of Thought for the Extended Mind: Cognitive Enhancement and the Constitution." *Wisconsin Law Review*: 1049.

Berners-Lee, T., Fischetti, M., *&* Foreword By-Dertouzos, M. L. (2000). *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor*. Harper Information.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American*, *284*(5), 28-37.

Bonjour, L. (1985). *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard University Press.

—— (2002). 'Internalism and Externalism', *Oxford Handbook of Epistemology*, (ed.) P. Moser, 234-64, Oxford: Oxford University Press.

Burge, T. (1986). 'Individualism and Psychology', *Philosophical Review* 95, 3–45.

———. (1993). 'Content Preservation', *The Philosophical Review* 102(4), 457-488.

Carbonell, V. (2013). 'What We Know and What We Owe', *Oxford Studies in Normative Ethics* 3.

Carter, J.A. (*Forthcoming*). 'Sosa on Knowledge, Judgment and Guessing', *Synthese*, 1-22.

Carter, J.A., Jarvis, B. *&* Rubin, K. (*Forthcoming*). 'Belief without Credence', *Synthese,* 1-29.

Carter, J. A., *&* Palermos, S. O. (2015). 'Active Externalism and Epistemic Internalism', *Erkenntnis 80*(4), 753-772.

———. Forthcoming. 'The Ethics of Extended Cognition: Is Having your Computer Compromised a Personal Assault?, *Journal of the American Philosophical Association.*

Chemero, A. (2009). *Radical Embodied Cognitive Science.* MIT press.

Chisholm, R. M. (1977). *Theory of Knowledge* (2nd ed.), Englewood Cliffs, NJ: Prentice-Hall.

Clark, A (2003) Natural-born Cyborgs: Minds, Technologies, and the Future of Human Intelligence (Oxford University Press, NY)

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

———. 2015. 'What 'Extended Me' Knows', *Synthese*, *192*(11), 3757-3775.

Clark, A., and Chalmers, D. (1998). 'The Extended Mind', *Analysis*: 7–19.

DeCew, J. (2013). 'Privacy', In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Fall 2013. http://plato.stanford.edu/archives/fall2013/entries/privacy/.

Froese, T., Gershenson, C., and Rosenblueth, D., A. (2013). 'The Dynamically Extended Mind'. http://arxiv.org/abs/1305.1958

Giere, R. (2002). 'Scientific Cognition as Distributed Cognition'. In *Cognitive Bases of Science*, eds. Peter Carruthers, Stephen Stitch and Michael Siegal, Cambridge: Cambridge University Press, 2002.

Giles, J. (2005). 'Internet Encyclopaedias Go Head to Head', *Nature*, *438*(7070), 900-901.

Greco, J. (1999). 'Agent Reliabilism', in *Philosophical Perspectives 13:* Epistemology (1999). James Tomberlin (ed.), Atascadero, CA: Ridgeview Press, pp. 273-296.

—— (2004). 'Knowledge As Credit For True Belief', in *Intellectual Virtue: Perspectives from Ethics and Epistemology*. M. DePaul & L. Zagzebski (eds.), Oxford: Oxford University Press.

—— (2007) 'The Nature of Ability and the Purpose of Knowledge', *Philosophical Issues* 17, pp. 57- 69.

—— (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press.

—— (2012). 'A (Different) Virtue Epistemology', *Philosophy and Phenomenological Research* 85 (1), 1-26.

Huemer, M. (2007). 'Compassionate Phenomenal Conservatism', *Philosophy and Phenomenological Research*, *74*(1), 30-55.

Hutchins, E. (1995). *Cognition in the Wild.* Cambridge: MIT Press.

Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.

Lynch, M. P. 2013. 'Privacy and the Threat to the Self', *The New York Times* (June 22).

http://opinionator.blogs.nytimes.com/2013/06/22/privacy-and-the-threat-to-the-self/.

———. 2014. 'Neuromedia, Extended Knowledge and Understanding', *Philosophical Issues* 24 (1): 299–313.

Menary, R. (ed.) (2010). *The Extended Mind*. Cambridge, MA: MIT Press.

Noë, A. (2004). *Action in Perception.* Cambridge, MA:MIT Press.

Noveck, B. S. (2007). Wikipedia and the Future of Legal Education. *J. Legal Educ.*, *57*, 3.

Pappas, G. (2014). 'Internalist vs. Externalist Conceptions of Epistemic Justification'. *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2014/entries/justep-intext/>.

Palermos, S.O. (2014a). 'Knowledge and Cognitive Integration', *Synthese* 191 (8): 1931–1951.

———. (2014b). 'Loops, Constitution, and Cognitive Extension', *Cognitive Systems Research* 27: 25–41.

———. (2015). Active externalism, virtue reliabilism and scientific knowledge. *Synthese*, *192*(9), 2955-2986.

——— (Forthcoming*a*). The Dynamics of Group Cognition. *Minds and Machines.*

——— (fprthcomingb). Social Machines: A Philosophical Engineering. *Phenomenology and the Cognitive Sciences.*

Palermos, S. O. & Pritchard, D. (Forthcoming). 'The Distribution of Epistemic Agency. In Socila Epistemology and Epistemic Agency: De-Centralizing Epistemic Agency, (Ed.)). Reider, (Rowman and Littlefield).

Plantinga, A. (1993*a*). *Warrant and Proper Function.* New York: Oxford University Press.

—— (1993*b*). *Warrant: The Current Debate*, Oxford: Oxford University Press.

Poston, T. (2008). 'Internalism and Externalism in Epistemology', *Internet Encyclopaedia of Philosophy*, (eds.) B. Dowden *&* J. Fieser, www.iep.utm.edu/int-ext/

Pritchard, D. (2010). 'Cognitive Ability and the Extended Cognition Thesis', *Synthese 175*: 133-151.

—— (2009). *Knowledge,* London: Palgrave Macmillan.

Pryor, J. (2001). 'Highlights of Recent Epistemology', *British Journal for the Philosophy of Science* 52, 95-124.

Pryor, J. (2004). 'What's wrong with Moore's Argument?', *Philosophical Issues*, *14*(1), 349-378.

Putnam, H. (1975). 'The Meaning of "Meaning"', *Minnesota Studies in the Philosophy of Science* 7: 131–193.

Simmons, A. J. (1993). *On the edge of Anarchy: Locke, Consent, and the Limits of Society*. Princeton: Princeton University Press.

Sosa, E. (1988).'Beyond Skepticism, to the Best of our Knowledge'. *Mind,* New Series, vol. 97, No.386, pp. 153-188

—— (1993). 'Proper Functionalism and Virtue Epistemology'. *Nous,* Vol. 27, No. 1, 51-65.

—— (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge,* Oxford: Clarendon Press.

—— (2015). *Judgment and Agency*, Oxford: Oxford University Press.

Steup, M. (1999). 'A Defense of Internalism', *The Theory of Knowledge: Classical and Contemporary Readings* (3rd Ed.), (ed.) L. Pojman, 310-21, Belmont, CA: Wadsworth.

Sutton, J. (2008). 'Between Individual and Collective Memory: Coordination, Interaction, Distribution'. *Social Research*, 75 (1), pp. 23-48.

Sutton, J., Barnier, A., Harris, C., Wilson, R. (2008). 'A conceptual and empirical framework for the social distribution of cognition: The case of memory'. *Cognitive Systems Research,* Issues 1-2, pp. 33–51.

Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., *&* Fergus, R. (2013). 'Intriguing Properties of Neural Networks', *arXiv preprint arXiv:1312.6199*.

Theiner, G. & Allen, C. & Goldstone, R. (2010). 'Recognizing Group Cognition'. *Cognitive Systems Research*, Vol. 11, Issue 4, pp. 378-395.

Theiner, G. & Allen, C. & Goldstone, R. (2010). 'Recognizing Group Cognition'. *Cognitive Systems Research*, Vol. 11, Issue 4, pp. 378-395.

Vallentyne, P. and van der Vossen, B. (2014). 'Libertarianism', *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2014/entries/libertarianism/>.

Wegner, D. M. (1987). 'Transactive Memory: A Contemporary Analysis of the Group Mind', In *Theories of Group Behavior* (pp. 185-208). Springer New York.

Wittes, B. and Chong, J. (2014). *Our Cyborg Future: Law and Policy Implications*. Brookings Centre for Technology Innovation.

Wilson, R. A., *&* Clark, A. (2005). 'How to Situate Cognition: Letting Nature Take its Course. *The Cambridge Handbook of Situated Cognition*.