Clowes, R. W., Smart, P. R., & Heersmink, R. (in press) The Ethics of the Extended Mind: Mental Privacy, Manipulation and Agency. In B. Beck, O. Friedrich & J.-H. Heinrichs (Eds.), *Neuro-ProsthEthics: Ethical Implications of Applied Situated Cognition*. Springer, Cham, Switzerland.

The Ethics of the Extended Mind: Mental Privacy, Manipulation and Agency

Robert W Clowes, Paul Smart & Richard Heersmink

Abstract

According to proponents of the extended mind, bio-external resources, such as a notebook or a smartphone, are candidate parts of the cognitive and mental machinery that realises cognitive states and processes. The present chapter discusses three areas of ethical concern associated with the extended mind, namely mental privacy, mental manipulation, and agency. We also examine the ethics of the extended mind from the standpoint of three general normative frameworks, namely, consequentialism, deontology, and virtue ethics.

Keywords: Ethics; Extended Mind; Extended Cognition; Ethical Parity; Mental Privacy; Mental Manipulation; Mental Agency; Responsibility

1. A New Ethical Landscape?

The extended mind hypothesis has been one of the most influential ideas originating in philosophy over the last 25 years. According to this hypothesis, artefacts, objects and other individuals may count as a constitutive part of a person's mind (Clark & Chalmers 1998; Clark 2008). There has been much debate about the metaphysics of the extended mind, but, until recently, the practical and normative consequences have been little explored. This is starting to change (e.g., Levy 2007; Heersmink 2017a, b; Heinrichs 2017, 2021; Carter & Palermos 2016; Clowes 2015). The extended mind hypothesis has changed the way we think about our relation to the local environment, and ethical issues are a plausible next step in its intellectual trajectory.

For those unfamiliar with the extended mind, we will briefly introduce the case of Otto (Clark & Chalmers 1998). Otto is a man afflicted by a deterioration in bio-mnemonic capabilities, incurred as the result of a mild form of dementia. As a coping strategy, Otto uses a notebook to aid him in remembering important information. Thus, when Otto is in New York and desires to visit the Museum of Modern Art (MoMA), he automatically consults his notebook and retrieves the information that MoMA is located on 53rd Street. According to Clark and Chalmers, the information in Otto's notebook plays more or less the same role in guiding Otto's thoughts and actions as does the information typically stored in biological memory. Given this, Clark and Chalmers suggest that we ought to regard the notebook (and its informational contents) as part of the supervenience base of Otto's dispositional beliefs. If the information had been retrieved from bio-memory, they suggest, then we would have little problem in regarding the bio-memory system as part of the supervenience base for Otto's beliefs (and thus a *bona fide* part of the machinery of his mind). Given this, however, it is hard to see why we ought to regard the notebook any differently. If both the notebook and bio-memory provide us with a suitable folk psychological grip over Otto's overt behaviour, then perhaps they both ought to be afforded equal cognitive status. That is to say, they both ought to be regarded as *bona fide* constituents of Otto's mind.

In support of such claims, Clark and Chalmers refer to a set of criteria that have come to be known as the trust+glue criteria.[1] In short, Clark and Chalmers claim that what makes the notebook part of Otto's mind is the fact that Otto has a certain relation to the notebook. What is crucial to the Otto case, Clark and Chalmers suggest, is that Otto has a high degree of trust in the notebook, he relies upon it, and it is readily accessible. When Otto desires to go to MoMA, he automatically consults the notebook, the relevant information is easily retrieved, and, upon accessing it, Otto automatically endorses it—he does not subject it to critical scrutiny in the way that we might treat information from a suspect news source.

In this chapter, we focus on exploring some of the ethical issues associated with the extended mind. We also reflect on how the extended mind—and the broader notion of cognitive extension[2]—might help us reframe new aspects of the ethical landscape that we inhabit. The idea of the extended mind may be particularly apposite to our historical moment, and its ethical implications especially useful to follow-through. It is thus worth briefly exploring why the concept of the extended mind has been so influential in recent times. There are arguably two main reasons for this.

First, much contemporary cognitive science has a strongly anti-Cartesian orientation that emphasises the need to understand cognition in its active, world-involving situated and embodied forms. So-called 4E cognitive science—emphasising the embodied, embedded, enactive, and extended nature of cognition—has directed its attention to the various ways in which cognition is actively produced through a series of embodied interactions with the proximal environment. One central, but often underestimated, element of this proximal environment (the local cognitive ecology) is the human-made world of artefacts, and it has become increasingly obvious that artefacts play a central role in human cognition (Donald 1991; Hutchins 1995; Clark 2008). This artefactual dependence naturally tends to evoke the concept of the extended mind.

Second, the early twenty-first century has seen an explosive growth of technologically-advanced artefacts—sometimes called "smart artefacts" or "smartefacts" (Carter and Palermos 2016). Such artefacts typically come equipped with networking capabilities that enable them to be connected to the Internet. In addition, they sometimes incorporate and present capabilities delivered as a result of recent advances in Artificial Intelligence (AI). As many millions of us are now constantly accompanied by their smart devices, the human impact of arguments for the extended mind become increasingly important. In particular, it is important to gauge the extent to which new technological devices may, just like Otto's notebook, become an intrinsic part of the human cognitive economy. This pushes us into new normative territory. How should we live with such devices? What should we require of them?

---

1 A number of criteria have been proposed to individuate cases of cognitive and mental extension. For reasons of space, we do not consider these additional criteria. See Heersmink (2015), for a review of some of the criteria that have been discussed in the literature.

2 A distinction is sometimes made between extended cognition and the extended mind, with the former centred on explanatory kinds relevant to cognitive science (e.g., extended problem-solving), and the latter centred on explanatory kinds relevant to folk psychology (e.g., dispositional belief). In the present paper, the term "cognitive extension" should be understood as referring to both extended cognition and the extended mind.

What should they do, and what should they not do? And what is the ethical impact of these new-found forms of bio-technological bonding?

In what follows, we begin by conceptualising the ethical significance of the distinction between embedded and extended cognition (Section 2). We then focus on three areas of ethical concern for the extended mind, namely mental privacy (Section 3), mental manipulation (Section 4), and agency and autonomy (Section 5). We conclude (Section 6) by examining the ethics of the extended mind from the standpoint of three general normative frameworks, namely, consequentialism, deontology, and virtue ethics.

2. Embedded and Extended Ethics

A commitment to the extended mind requires a radical shift in the way we think about bio-external resources. According to proponents of the extended mind, a bio-external resource (such as a notebook or smartphone) may not just be a causally-relevant feature of our extra-organismic environment; it may also be a constituent of our cognitive and mental machinery. There is an important distinction here between the notions of causal and constitutive relevance. Proponents of the extended mind, it should be clear, are wedded to the notion of constitutive relevance—they insist that bio-external resources can, on occasion, count as the *bona fide* constituents of mental states and cognitive processes. The alternative view comes in the form of what is called embedded cognition (Rupert 2004). According to the proponents of embedded cognition, bio-external resources ought not to be seen as the constituents of the mind; instead, such resources exert a merely causal influence on cognitive processes that operate behind the traditional metabolic firewalls of skin and skull. As noted by Wheeler (2019), the contrast between extended and embedded cognition is one that is typically understood with respect to the difference between constitutional and causal claims:

> […] in cases of extended cognition, the machinery of mind stretches beyond the skull and skin, in the sense that certain external elements are, like an individual's neurons, genuine constituents of the material realizers of that individual's cognitive states and processes […] By contrast, in cases of what is now often called embedded cognition, the machinery of mind remains internal, but the performance of that inner mental machinery is causally scaffolded in significant ways by certain external factors. (Wheeler 2019, p. 861)

There are reasons to think that this causal/constitutive distinction has a bearing on ethical debates and discussions. This significance was first noted by Clark and Chalmers (1998) as part of their seminal treatment of the extended mind. In particular, Clark and Chalmers (1998, p. 18) note that "in some cases interfering with someone's environment will have the same moral significance as interfering with their person." Such sentiments are reflected in the work of a number of subsequent authors. Carter and Palermos (2016), for example, suggest that an extended mind resource ought to have the same ethical and moral entitlements as those associated with a more conventional mind part (e.g., a part of an individual's biological brain). The deliberate destruction of a cognitively incorporated smartphone, they suggest, could be deemed as ethically and morally abhorrent as a form of neurological assault.

There may be ethically-motivated reasons to countenance an embedded versus extended approach. One reason to adopt an embedded approach is that it may be easier to assign

ethical responsibilities regarding the use (or misuse) of specific resources. Consider, for instance, a case where a recruitment consultant, let's call her Janet, makes frequent use of an AI-driven system to evaluate the candidacy of prospective employees. Let us imagine Janet has come to rely on the system in a way that satisfies the criteria for cognitive extension. Accordingly, the AI system counts as part of her cognitive system, or part of her mind. Now let us imagine that the AI system is found to support biased decisions that discriminate between candidates based on some protected characteristic (e.g., gender or ethnicity). At this point we confront an important question: Who is responsible for the prejudicial employment decisions?

From an embedded standpoint, the ethical spotlight seems to fall on the developer of the AI system. After all, Janet's decisions were made thanks to the influence of a relatively opaque piece of software, one that Janet used (as a mere tool) in the context of a specific cognitive task.

From an extended standpoint, however, things look rather different: the recruitment decision was reached as a result of the whirrings and grindings of a cognitive system that includes (as proper parts) both Janet and the AI system. Indeed, the AI system forms part of Janet's cognitive machinery, just as would be the case if the functionality of the AI system had been realized by the operation of Janet's brain-based neural circuits. This looks to be important given that we are typically held personally responsible for our decisions, i.e., the outcomes of our cognitive apparatus. Since the biased decisions were made as a result of Janet's cognitive apparatus, it looks as though she ought to be held responsible for such decisions. This is, after all, a common feature of active externalist debates. It is thus typically assumed that the subject of cognitive extension—the extended subject—is in some way responsible for the outputs delivered by an extended cognitive circuit (e.g., Roberts 2012). As a means of reinforcing this particular point, let us consider the case of Anna, an architect who relies on Computer-Aided Design (CAD) software to formulate architectural design solutions (see Clowes 2020). Do we want to say that such solutions are to be credited to the creator of the CAD software rather than Anna? This seems unlikely, since the design solution owes its existence to the facility that Anna has with the relevant software package. In addition, there is surely a sense in which Anna deserves much of the credit for controlling and coordinating the time-variant flow of information around a causally-interacting nexus of material objects in such a way as to forge the path to a successful design solution. The provision of the solution is a cognitive achievement of sorts, and in recognizing who is responsible for this achievement, the spotlight seems to fall on Anna.

The issues here are subtle, but it quickly becomes clear that the choice between extended and embedded seems to have important consequences for how we understand the ethics of bio-technological interaction. At an individual level, there are questions regarding what tools and technologies we ought to rely on, and what features or properties they ought to possess. Such questions are perhaps common to both extended and embedded perspectives, but they are arguably accentuated when the focus shifts to an extended perspective. From an embedded perspective, a bio-external resource remains separate from us. But once we shift to an extended perspective, those same resources become a part of us.

3. Mental Privacy

A number of theorists have written about the privacy-related implications of the extended mind (Smart et al. 2017; Carter et al. 2018; Carter 2021; Palermos 2023). The core concern, here, relates to the accessibility of bio-external resources. If, for example, we accept that Otto's notebook counts as part of his extended mind, then it seems that we might be able to 'read' Otto's mind simply by reading his notebook. If Otto is asleep, for example, we might surreptitiously access his notebook and see the inscription pertaining to the location of MoMA. Upon reading this, we can infer that Otto believes (in a dispositional sense) that MoMA is located on 53rd Street: if Otto desires to visit MoMA, then we know where Otto will go.

Claims about the extended mind therefore lead to a concern about mental privacy and the possibility of enhanced forms of cognitive/mental surveillance. If bio-external resources can count as part of the machinery of the mind, then it seems perfectly possible for some third party to scrutinize an individual's thoughts simply by accessing the bio-external resource.

Such concerns are particularly prominent when it comes to our contemporary reliance on network-enabled devices (e.g., smartphones) and online content. Suppose, for example, that we introduce a technological upgrade to the Otto case, whereby we replace the notebook with a portable computing device, and we relocate the notebook's contents to an online (cloud-based) data store. Call this the Accessible Otto case. In view of the technological upgrades, it seems that anyone with access to the online data store must be suitably poised to remotely monitor Otto's mind. They can thus monitor whatever entries are made to Otto's online data store, and they can gain access to Otto's thoughts simply by reading its contents. None of this requires them to wait until Otto is asleep. Nor do they need to be in the physical presence of Otto. In principle, the contents of Otto's mind are accessible to a number of different social actors, ranging from corporate entities, individual hackers, and government bodies. The old idea of privileged access to one's own mental states, so important to Descartes, is here clearly placed under severe stress.

A commitment to the extended mind thus entails a suite of privacy-related concerns, many of which go beyond the traditional ones that animate debates into data privacy. In response to this, it has been suggested that we ought to introduce tighter controls on who has access to extended mind resources. Orestis Palermos (2023), for example, suggests that access to the contents of an extended mind ought to be restricted to the individual whose mind is being extended. Thus, in the Accessible Otto case, the only individual who ought to be able to access the online data store is the biological individual we know as Otto; all other forms of access ought to be prohibited. For the sake of convenience, let us refer to this as the For My Eyes Only (FMEO) principle—the principle that only the subject of cognitive extension ought to be able to view (or access) the information that serves as the bio-external basis of *their* extended beliefs.

At first sight, the FMEO principle looks to be a perfectly reasonable response to concerns regarding mental privacy. There are, however, a number of problems with the principle. One problem relates to situations where the contents of an extended mind are shared across multiple individuals. Consider, for the sake of example, a concrete attempt to build a technological system that supports extended beliefs and knowledge. The system, described by Smart (2021), aims to furnish users with information about works of art. The functionality of the system relies on an online database called DBpedia, which is a machine-readable

version of Wikipedia. Crucially, this database is a common resource that is shared by multiple users. The DBpedia resource, in this scenario, forms part of what we might call the 'cognitive commons'—it is a publicly-accessible resource that (at least in principle) forms part of multiple extended minds. Let us assume that DBpedia meets the conditions for at least one individual to count as part of his or her extended mind.

It is hard to see what a strict commitment to the FMEO principle would buy us in this sort of situation. The relevant body of information—that contained in DBpedia—is already in the public domain, so the privatisation of this information (e.g., the creation of a privately-accessible clone of the DBpedia database) doesn't really amount to much—we can simply refer to the publicly-available version of the DBpedia database to tell us what the privatised version contains. Perhaps, more importantly, however, privatisation doesn't really serve the cognitive and epistemic interests of the relevant user community. As noted above, DBpedia is a machine-readable version of Wikipedia, and part of the reason why Wikipedia is such a potent epistemic resource is that it is subject to public scrutiny. This means that any factual errors can be quickly identified and remedied, thereby ensuring the reliability of the information content. In this case, it is the public accessibility of the resource—the fact that it is visible to *many* eyes—that helps to ensure its epistemic integrity. To privatise the resource (or at least to privatise access to the resource) is to jeopardize the very thing that makes the resource a fitting target for bio-technological bonding.

The FMEO principle entails a further problem, although the problem is one that is relevant to other areas of ethical concern. The problem is that proponents of the extended mind typically assume that the constituents of an extended mind (both physical and informational) are deserving of special ethical treatment and attention. Carter and Palermos (2016), for example, note that if a bio-external resource (e.g., a smartphone) counts as part of an extended mind, then it ought to be afforded certain ethical and legal protections. It would not be appropriate, they suggest, for some external actor to gain unauthorised access to the smartphone, because this would amount to a violation of mental privacy. Similarly, any attempt to destroy the smartphone now looks to be morally abhorrent. If the smartphone counts as part of an individual's extended mind, then its metaphysical status is on a par with the neural circuits that comprise the individual's biological brain.

At first sight all this sounds eminently plausible. The problem is that the status of an extended mind resource, i.e., a resource that forms part of an extended mind, presents us with an epistemic problem. How do we know that any given resource (e.g., a smartphone, a notebook, or the contents of an online data store) ought to be seen as a constituent of an extended mind? In philosophical circles, this question lies at the heart of a long-standing debate regarding the evaluation of putative cases of cognitive extension. In particular, philosophers are concerned with the criteria that distinguish genuine cases of cognitive extension from those of the merely embedded variety. For present purposes, we can overlook the details of this debate; what is important is simply the fact that cognitive extension comes with an epistemic challenge: In order to know that a resource qualifies as a *bona fide* constituent of an individual's mind, we need to understand something about the role the resource plays in the cognitive and mental life of a given individual. According to Clark and Chalmers (1998), for example, we need to know that the resource is subject to typical invocation, that it is

accessible to an individual, and that the individual does not treat the informational deliverances of the resource with suspicion or critical scrutiny.

This leads to a potential problem for the proponent of the FMEO principle. Suppose you are someone who works for a cloud-based computing company. You wholeheartedly endorse the FMEO principle, and thus you have no interest in accessing the contents of someone's extended mind. The problem is that in order for you to act in an ethically appropriate manner, it seems that you need to know which online resources count as part of someone's extended mind. But in order to make this discrimination, it seems that you must first engage in a degree of monitoring and scrutiny to learn about the role that specific online resources (e.g., bodies of information) play in the cognitive and mental life of a human individual. This, of course, raises its own privacy-related concerns.

In short, then, the worry about the FMEO principle is that it does not really attenuate privacy-related concerns. In order to treat bio-external resources in a manner that respects privacy-related (and other ethical) constraints, we need to know that those resources count as *bona fide* constituents of an extended mind. But in order to know this, we are in danger of violating the very thing that the FMEO principle was intended to safeguard against. In particular, the worry is that we create an ethical obligation to subject individuals to ever-greater levels of surveillance and scrutiny, with the aim of acquiring knowledge about their extra-cranial cognitive architecture. This knowledge may very well serve as the basis for decisions and actions pertaining to the ethical treatment of extended mind resources (e.g., not accessing the contents of an online data store). But the problem is that the mere acquisition of this knowledge comes with its own attendant set of ethical concerns. In a privacy-related context, the worry is that one sort of normative constraint (i.e., do not access the contents of another person's extended mind) serves as the ethical mandate for actions that (from a privacy-related perspective) are perhaps no less egregious than the actions the FMEO principle seeks to prohibit.

Despite its ostensible shortcomings, there is clearly something right about the FMEO principle. It is appropriate, it seems, to expect a degree of mental privacy, if only because this speaks to the libertarian emphasis attached to freedom of thought (see Blitz 2010). The question, then, is whether some variant of the FMEO principle might be sustained by interventions that do not jeopardize the privacy-related rights of extended subjects. In response to this question, it is worth noting that privacy-related concerns lie at the heart of a number of scientific, engineering, and policy-making efforts.[3] There has, of course, been

3 Two further strands of research that may be relevant to issues of mental privacy are directed to issues of data access and sharing. The first of these centres on what are called personal (online) data stores (Mansour et al. 2016; Van Kleek and O'Hara 2014). These are data repositories that are intended primarily for personal access and use, although they sometimes allow for limited forms of data sharing. A second strand of research relates to the notion of trustworthy data institutions. Work in this area falls under a number of headings, such as data trusts, data safe havens, data foundations, and trusted research environments (Boniface et al. 2020; O'Hara 2020; Patel et al. 2022). What is common to all these locutions is the idea that a trusted third-party (an individual or organization) is assigned to protect the interests of a data subject, where the notion of "interest" extends to include matters of data privacy. While this approach is typically discussed in relation to bodies of personal data, there is no reason why the same sort of approach could not be applied

plenty of research into biometric identification and cryptographic protocols, and much of this speaks to the general concern with individualized access to specific bodies of data. There has also been considerable work into data anonymisation techniques (e.g., Elliot et al. 2018). Anonymisation looks to be important for mental privacy, insofar as privacy-related concerns are intimately connected to matters of identification.

Yet, it is doubtful that anonymization can resolve all the ethical issues in this area. Consider a host of individuals using a collective online system to store large amounts of annotated data. Now let us suppose this data meets the trust + glue conditions for at least some of its users and therefore counts as part of those users' extended minds (and thus as part of their store of dispositional beliefs). The company that stores the data, let's call them Mass Analytica, promises as part of their terms of use to anonymise all data collected. But Mass Analytica stores the data and makes services available for commercial reasons. They intend to use the stored data for a number of purposes including the training of a deep learning system for the prediction of consumer choices, and especially a system to predict the susceptibility of types of voters to political influence. Mass Analytica further intends to use the data harvested in this way to influence the outcome of an election by influencing (or manipulating) some of the very users whose data has been harvested.

One problem with such "mind-reading" AI is that even anonymised data can still be used to train algorithms for an open-ended number of purposes. Such cases of the "dual use" of data are known to be ethically problematic (see Becker et al. 2023).[4] Exactly what information from someone's mind should be available for an open-ended use for data mining and prediction purposes? Subjects would be unlikely to knowingly consent to their brain scans being used in commercial applications that might later be used to manipulate them (we will discuss manipulation further in the next section). Should this be different in cases where the mental resources being so used are parts of their extended minds? If dual use scenarios are already known to be problematic with AI more generally, how much more so might they be when the data and algorithms might also be considered to be part of an individual's mind? In this way, concerns about the privacy of extended minds shade into our next area of concern which is mental manipulation.

4. Mental Manipulation

As we have seen, privacy-related concerns are tied to issues of accessibility. The fact that a bio-external resource (e.g., notebook or smartphone) lies external to the corporeal boundary of a human individual means that it can be accessed in such a way that does not breach bodily borders. This accessibility comes with an additional worry, however. The worry is that an individual who has access to the contents of an extended mind resource might be in a position to interfere with those contents. To help us understand this worry, consider that if one were

---

to address some of the ethical concerns (not just those associated with mental privacy) that arise in respect of the extended mind.

4 Several attempts to develop controls on third-party activities to ensure their compatibility with the desires, wishes, and ethical convictions of the (anonymised) data subject have been developed. There is a potentially fruitful link here with work into policy-aware computing (Weitzner et al. 2006), purpose-based data access protocols (Kraska et al. 2019), and regulatory compliance modelling (Taylor et al. 2021).

to gain access to Otto's notebook—perhaps while Otto was sleeping—then one could surreptitiously manipulate the information contained in the notebook. Someone of malicious intent could, for example, modify the entry relating to MoMA's location. Instead of the notebook reading that MoMA is on 53rd Street, the malicious intruder might modify the entry to read that MoMA is on 43rd Street. The upshot is that Otto will no longer go to the museum's actual location when he desires to visit MoMA. Instead, Otto will go to 43rd Street. It seems, then, that by modifying the notebook, the intruder has succeeded in effecting a form of mental manipulation. In particular, the intruder seems to have succeeded in changing one of Otto's dispositional beliefs, specifically that pertaining to the location of MoMA. Prior to the manipulation Otto believed that MoMA was on 53rd Street, but subsequent to the manipulation he believes that MoMA is on 43rd Street (or, perhaps, more worryingly, he believes whatever the intruder wants him to believe!).

A recent attempt to explore issues of mental manipulation from an extended mind perspective stems from Carter (2021). Carter identifies two forms of mental manipulation that might arise in an extended mind context. These concern the acquisition of new beliefs (acquisition manipulation) and the erasure/deletion of existing beliefs (eradication manipulation). In some situations, of course, there is nothing wrong with the acquisition or eradication of beliefs, but in situations where the relevant manipulation is made in a covert or clandestine manner—as in the aforementioned sleeping Otto case—then attempts at mental manipulation are apt to lead to ethical problems.

At first sight, the ethics of mental manipulation look to be largely uncontroversial. Quite plausibly, we would not welcome covert forms of mental manipulation that targeted our brain-based body of beliefs, and the same seems to be true of beliefs that stem from the operation of an extended cognitive circuit. Accordingly, it seems that we might be able to posit an ethical principle to the effect that no intervention may be made in an individual's mental economy (extended or otherwise) without the express prior consent of the individual concerned.

While some variant of this ethical principle might be made to work, it is important to note that a blanket ban on mental manipulation confronts a number of problems. To help us understand this, consider a state-of-affairs in which an individual relies on an app to deliver accurate information. In particular, let's imagine the sort of scenario discussed by Clark (2007) where an individual is equipped with an augmented reality display that provides easy and efficient access to facts and figures about women's basketball. By participating in this extended cognitive/epistemic system, we can assume that the human individual expects the app to function in a reliable manner. That is to say, it is not unreasonable to think that the user expects the app to provide them with factually correct information. Indeed, the very reason this particular form of bio-technological bonding exists is because the individual expects the app to enhance their epistemic standing relevant to the focal epistemic domain (i.e., the body of facts pertaining to women's basketball).

Now let's suppose that the manufacturer of the app discovers a bug in the software that leads the app to deliver the wrong result in certain situations. Given the expectations of the user, we might expect the manufacturer to fix the fault, so as to ensure the veracity of the user's beliefs. But by implementing this update, the manufacturer has arguably engaged in a form of mental manipulation—they have, after all, modified the informational deliverances of the

app such that an erstwhile incorrect belief now succeeds in tracking the factive nature of reality.

Should this particular form of mental manipulation be regarded as ethically problematic? It is hard to see why. After all, the user may very well expect the software manufacturer to implement these sorts of updates, so as to ensure their epistemic integrity. In this sense, the manufacturer's failure to intervene in the extended mental economy may seem to be just as ethically problematic as does the more malign forms of manipulation observed in the sleeping Otto case.

In response to this, it might be suggested that the manufacturer has an ethical duty to notify or alert the user as to the nature of the change, and this notification perhaps ought to occur prior to the change being made. In practice, however, it is hard to see what difference this would make to either the user or manufacturer. Given the epistemically oriented concerns of the user, it is, in particular, hard to see why the user would object to the update. Perhaps more worryingly, however, the appeal to alerts and notifications threatens to aggravate the ethical concerns we explored in the previous section. In order for the manufacturer to alert the user, and perhaps seek their explicit consent, the manufacturer needs to know (at a minimum) how to contact the user. But does the manufacturer really need to know this information? And doesn't the possession of this knowledge raise precisely the sort of worries that informed the discussion of mental privacy? Presumably, the best way to safeguard the individual's mental privacy in this scenario is for the manufacturer to *not* know anything about the subject of mental extension. That is to say, the best way to protect the mental privacy of the extended subject is for the extended subject to remain anonymous. If, however, we are insisting that each individual should be contacted prior to (or even subsequent to) an update, then it is hard to see how this could be the case.

The ethics of mental manipulation are also complicated by a socially-inflected form of cognitive extension, known as socially-extended cognition (e.g., Clark and Chalmers 1998). In socially-extended cognition, another human individual plays the role that is typically attributed to a technological or artefactual resource. Such forms of cognitive extension challenge the idea that no form of covert mental manipulation is permissible in extended mind scenarios, because such an ethical constraint threatens to jeopardize the mental autonomy and liberty of another human individual. We cannot endorse the idea that individuals have a right to mental autonomy (conceived of as the right to alter their beliefs in a manner they see fit) while also endorsing the idea no form of mental manipulation is permissible in extended mind scenarios. Such commitments might be made to work in situations where an extended mind is constituted by non-human artefacts, but it cannot be made to work in situations where another human individual functions as part of a socially-extended mind. To embrace such commitments in a socially-extended context introduces an inevitable tension: it trades the rights of one individual against the rights of another. In particular, to assume that one individual cannot change their mind on the grounds that it would constitute a form of mental manipulation is to challenge an individual's entitlement to freedom of thought, and such freedoms are deemed important, even to those who endorse the extended mind hypothesis (see Blitz 2010).

5. Agency and Autonomy

The reference to mental autonomy serves as a convenient segue into our third area of ethical concern. This relates to the impact of cognitive extension on our mental autonomy, i.e., our capacity to influence the course of our cognitive and mental evolution and thus determine the nature of who and what we are. Such concerns are particularly prominent in situations involving so-called AI extenders (Vold and Hernández-Orallo 2022). According to Vold and Hernández-Orallo:

> An AI extender is a cognitive extender that is "fueled" by AI. This means that some AI technology is directly responsible for the cognitive capability that the extender is able to deploy, in conjunction with its user. (p. 183)

In essence, then, an AI extender is an incorporated resource (a physical constituent of an extended cognitive circuit or extended mind) whose activities are, at least in part, governed by the operation of AI algorithms.

One of the issues raised by AI extenders is their capacity to anticipate user information requirements via the monitoring of user behaviour. Call this capacity proactive information retrieval (PIR). From an extended mind perspective, PIR is both a boon and a burden. It is a boon in the sense that it automates some of the activities that might otherwise need to be performed by the human individual, and, in this sense, it reduces the costs (both temporal and energetic) associated with the retrieval of (e.g.) belief-relevant information. Consider, for example, that if Otto was equipped with an AI extender that was able to anticipate his information requirements, then there would be no need for him to retrieve the information himself. Rather than consulting a notebook, the AI extender would simply retrieve the relevant information on Otto's behalf and present it to his perceptual apparatus. Accordingly, whenever Otto desired to go to MoMA, the information about MoMA's location would be available to him, just as it would if he had retrieved that information from bio-memory.

There is, however, a downside to all this. The worry is that by obviating the need for individuals to retrieve their own information, PIR may work to undermine the mental autonomy of the individual. To help us understand the nature of this worry, let us suppose you are equipped with an AI extender—one that is capable of anticipating your desires, based, perhaps, on a detailed record of your past behaviour. One day you decide to go on a diet and relinquish your habit of ordering takeaway meals. Unfortunately, your AI extender seems to have other plans. You now find yourself bombarded with reminders of your former lifestyle—details of special takeaway offers, enticing images of a now forbidden mambo chicken dish, and perhaps most unnervingly of all, an unexpected pizza delivery. It may be that in time your AI extender will adjust, in the interim, however, the AI system is (at best) an annoyance. And inasmuch as you succumb to the persistent barrage of culinary temptations, you may find it difficult to forsake your former self and become someone new.

Such concerns are not limited to the realm of AI extenders; they also apply to a variety of "smart" technologies, especially those that engage in user profiling and personalised display. Delacroix and Veale (2020, p. 6), for instance, worry that smart technologies may "undermine our capacity to develop and maintain an integral sense of self". Rather than helping us discover who we are, or what we may become, Delacroix and Veale worry that smart technologies may lead to self-fulfilling prophecies, ones in which we are progressively coaxed

into becoming the sort of individual that a technology perceives us to be (and/or what the technology's corporate sponsors want us to be).

Yet not all self-tracking systems, even those that can be considered AI extenders need impugn individual autonomy or agency.[5] Consider, for example, the self-tracking systems used by millions of individuals to collect data across a broad spectrum of physiological and behavioural parameters. With the rise of wearable devices such as the ubiquitous Fitbit and self-tracking apps, many millions of individuals are already tracking information pertaining to physical activity, menstrual cycles, sleep patterns, and caloric intake. Self-tracking systems—as exemplified by those featuring as part of the Quantified Self movement—are not just used to track information about one's body and activities; they are also used to control and modify one's own behaviour, as *self-shaping systems*. Such self-shaping often occurs via the use of data visualization dashboards in apps, the setting of goals, and the tracking of one's progress in respect of those goals. The data delivered by such systems is not merely used to support knowledge about the self ("self-knowledge through numbers," as per the motto of the quantified self movement); it is also used to effect changes in oneself—to not just confirm an image of who and what we are but to support the data-driven journey to a new self: an image of the person we may yet become.

Such self-shaping can be strongly linked to the practice of human agency. According to the philosopher Michael Bratman (2000), human agency is associated with a closely integrated set of cognitive skills that includes the capacity for prospection, planning, reflection, and self-regulation. Insofar as these capacities are amenable to cognitive extension, then our bio-technological mergers may serve as the material roots for a revised understanding of human agency. The point here is that not all technological systems need to be seen as a threat to human agency, in the sense that they undermine, limit, or erode human agency; instead, from an extended perspective, we can regard technologies as the potential constituents of extended cognitive circuits that are, themselves, the physical mechanisms that make our distinctively human agency the thing it is.

In one sense, then, we have no reason to regard self-tracking technologies or AI extenders as *necessarily* undermining human agency. Such systems can become parts of the extended mechanisms of human agency. On the other hand, the appeal to cognitive extension does not eliminate all the ethical concerns in this area. Consider how a person may come to rely on a self-tracking system to support their self-shaping efforts. From the user's perspective, the functionality of the system may look straightforward, but, behind the scenes, there may be much that the user is not aware of. Many wearable and app-based systems are designed to be highly "user-friendly" and therefore transparent-in-use (see Wheeler 2019). But the same

5 While Vold and Hernández-Orallo (2022) are correct to draw attention to the issues raised by AI extenders, it is important that we do not overstate the ethical implications of technologies that come equipped with some form of user profiling, user personalisation, or PIR-related capability. Not all AI extenders are committed to the modelling and monitoring of user behaviour (see, for example, Smart 2021), so it would be a mistake to tar all AI systems with the same ethical brush.

technologies are often rather reflectively opaque,[6] yielding little information about how they work, or what hidden biases they may build in (see Andrada et al. in press; Clowes 2020). As we come to rely upon such systems for making an ever-wider range of decisions on our behalf, they can often obscure the motivations and goals of their creators. Insofar as this impedes our abilities to anticipate the changes wrought by technologies, then human agency (in Bratman's sense) may be undermined.[7]

Towards the end of their discussion on AI extenders, Vold and Hernández-Orallo (2022) comment on the responsibilities of manufacturers when it comes to the design and maintenance of AI systems. In particular they note that:

> [...] from the point of view of cognitive extensions, the manufacturer must understand that the software and the hardware become part of the mind, so no updates, discontinuations or access to the data can be done without informed consent. Under a strict interpretation of the EM [extended mind] thesis, modifying an AI extender should be compared to modifying the brain. (Vold and Hernández-Orallo 2022, p. 34)

This point is made in relation to AI extenders, but the concern regarding updates and the ensuing shift in functionality is one that is applicable to a broad spectrum of technological devices, including those used for self-tracking and self-shaping purposes as just discussed.

One of the worries raised by software updates is that they potentially disrupt the local cognitive ecology of a human individual. To help us appreciate this concern, imagine that you come to rely on the functionality of a particular app for the purpose of performing a cognitive task. The app, let's suppose, is quite complex, and it takes you a number of weeks to become accustomed to its functionality. On first using the app, it is doubtful whether it can be considered a constituent of your cognitive/mental machinery, for cognitive extension is typically assumed to require a certain facility with bio-external resources. Over time, however, you become accustomed to using the app, and the app transitions from a mere tool to a *bona fide* mind part.

Now let's suppose that the designers of the app implement an upgraded version of the app, one which preserves the original functionality, but which alters the way certain computational routines are accessed. The installation of this upgrade, it should be clear, threatens to undermine your proficiency with the app, thereby leading to (at least a temporary) degradation in performance. The worry, then, is that software updates, as well as other externally-induced shifts in the local cognitive ecology of a human individual, can lead to (at least) a temporary disruption or degradation in cognitive performance. Modern digital devices afford plenty of opportunities for us to form extended cognitive circuits, but such circuits are often 'fragile' and transient constructions. In an era of frequent software updates, short product cycles, and innovation pressures, our cognitive circuits are seldom immutable.

---

6 The notion of reflective transparency refers to our capacity to 'see into' the mechanisms that realise the functionality of some system, thereby contributing to our understanding of how it works across a range of actual and counterfactual circumstances. Reflective opacity refers to the absence of this insight or understanding (see Andrada et al, in press, for more details).

7 For more on the ethical implications raised by transparent technology, see Wheeler (2019).

Existing circuits are sometimes undermined as a result of new innovations, or they may be destabilised as the result of shifts in government policy. In the worst case, an extended cognitive circuit may simply be obliterated as the result of obsolescence. The device that sustains one's cognitive endeavours in the here and now is unlikely to do so a decade hence.

Some insight into the hazards of changes to the local cognitive ecology is provided by Clowes (2020). We discussed earlier the case of Anna the architect whose professional abilities are tied to her facility with a CAD software package. Clowes suggests that the architect's sense of her own cognitive capabilities—her abilities to think, imagine, and solve architectural design problems—may be highly dependent on the nature of her interaction with the CAD software. If the architect's access to the software was suddenly curtailed, she might lose access to vital parts of herself, thereby compromising her sense of who and what she is. But even if the software were to radically change, then the architect's ability to use the software might be undermined. In the worst case, such a scenario might lead to a profound re-evaluation of the architect's own abilities. Prior to the upgrade the architect may have credited herself with the possession of certain cognitive abilities. But, subsequent to the upgrade, these sorts of self-related characterizations are apt to be called into question.

The moral, it seems, is that technology developers ought to be cautious about introducing changes to resources that may, in principle at least, count as the constituents of an individual's extended mind. At the same time, the ethical issues in this area are not straightforward. One problem relates to the ever-changing nature of the wider environment in which our technological devices are situated. Consider that many technological devices rely on what we might call distributed functionality. That is to say, their processing routines are realised by information processing circuits that reach out beyond the borders of the device to include all manner of external assets, including those situated in the online realm. The operation of a smartphone app, for example, may rely on access to the Internet for the purpose of retrieving certain bodies of data, or it may delegate certain computational routines to a remotely-situated Web service. Such forms of interconnectivity and interdependence make updates difficult to avoid, since any change in the wider ecology of computational assets will necessitate some sort of modification to the app's code base. The need for consent in such scenarios is something of a moot point, for if the update is refused, the app may no longer continue to function.

A similar sort of issue arises in respect of changes to the cyber-security environment. The point here is that apps and devices are prone to various forms of hacking and malign intervention, and updates are often required to guard against these. In such situations, the ethical onus is arguably on the manufacturer to implement and disseminate the update. Indeed, the failure to implement the update, and thus preserve existing cognitive circuits, seems to raise just as many ethical questions as does the attempt to impose an unsolicited update on an unsuspecting user.[8]

---

8 Note that the issue here is not one of practicality. That is to say, the problem of deploying updates to safeguard against a cyber-security vulnerability is not to be understood in terms of the cost or difficulty of deploying the update. The concern is more about whether or not the update *ought* to be deployed. There is no straightforward ethical response here. If an update is required to mitigate the risk of a cyber-attack, then a company might be seen to have an ethical obligation

6. Normative Framings for Extended (and Embedded) Ethics

In this paper, we have considered the ethical significance of the distinction between the embedded and extended cognition theses. We then focused on three specific ethical issues associated with the extended mind, namely mental privacy, mental manipulation, and issues to do with agency and autonomy. In this concluding section, we aim to situate these ethical issues in the context of more general normative frameworks, with the aim of determining what further guidance they can provide in respect of the ethics of the extended mind. Of particular interest is the extent to which these frameworks can help to resolve some of the dilemmas that were encountered in previous sections. In Section 3, for example, we saw how a commitment to the FMEO principle might in some circumstances exacerbate privacy-related concerns, rather than resolving them. Then, in Section 4, we encountered the idea that a blanket ban on any form of mental manipulation might conflict with the ethical obligation to ensure the veracity of information. Finally, in Section 5, we saw that companies might have an ethical duty to introduce software updates to (e.g.) safeguard against a cyber-security attack, despite the fact that such updates may simultaneously destabilize pre-established cognitive circuits.

In the present section, we survey three approaches to the extended mind—glossed as the consequentialist, deontological, and virtue ethical approaches—in the hope that these might offer guidance in respect of these dilemmas.

Consequentialism is the view that we ought to assess the ethical propriety of any action, intervention, or policy based on its implications for the persons involved (Mill 1891; for a more general overview, see Sinnott-Armstrong 2021). The obvious advantage of this framework is that it provides a way of factoring the interests of diverse stakeholders into our ethical deliberations. This looks to be particularly important in situations where we encounter competing or conflicting interests. Consider, for example, a situation where a technology company is considering the roll-out of a new software update. Such an update may work to the overall cognitive and epistemic benefit of some users, but it may also harm the cognitive and epistemic wherewithal of others. Imagine a case where a company is developing an app that is intended to remedy a deficit in autobiographical memory incurred as the result of Alzheimer's disease. As part of the testing phase, the app is trialled with a number of users, and these users come to rely on the app's functionality. Later, however, the company decides to change the functionality of the app, so that it can be deployed on a number of different devices, potentially benefiting millions of users. The problem is that, subsequent to the testing phase, the company can no longer afford to maintain the prototype app. Given the potential benefit, the consequentialist approach might suggest that the cognitive interests of those early users could be sacrificed for the sake of benefiting millions of others.

Consequentialism might help to frame some of these questions—where the benefits to the many outweigh the harms to the few. The questions faced by consequentialism are, as ever, how to justify difficult solutions that may genuinely benefit some individuals at the expense

---

to introduce the update. To fail to do this is to leave the individual user susceptible to malign intervention. But to introduce the update entails its own ethical problems (e.g., the destabilization of the local cognitive ecology). The point here is that there is no simple answer, either way: both action and inaction are ethically problematic.

of others. In an extended mind context, some of the decisions advocated by a consequentialist calculus may result in serious cognitive harm to one or more individuals.[9] For many, such implications will be unacceptable.

An alternative normative framework, the deontological approach, focuses on individual rights and duties where certain actions are held to be morally required, forbidden, or permitted (Alexander & Moore 2021). Deontological approaches do not seek to judge the morality of actions, interventions, or policies, based on consequentialist considerations. Instead, what makes a choice or action right is its conformity with a moral norm such as, for example, a respect for people's privacy or autonomy. A deontological approach to the extended mind might hold that any form of access or interference with an extended mind resource is ethically impermissible, and that this impermissibility obtains under any circumstance. In the case we just considered—the one relating to a pre-deployment shift in a dementia app's functionality— the company would be prohibited from making a change on the grounds that it would negatively affect the cognitive integrity of the individuals recruited during the test phase. The deletion of online data, a software update, or the withdrawal of rights to use a given software system, might similarly affect the cognitive wherewithal of a given individual. Accordingly, such actions would also be prohibited, from a deontological standpoint. Such potential cognitive damage to an individual might compel other deontologically influenced actors to take stringent steps to protect that individual's cognitive autonomy, even if those steps implied great inconvenience, cost, or meant curtailing or prohibiting certain sorts of technology development.

One criticism of the deontological approach is that it may be insufficiently flexible in the face of technological change. Shannon Vallor (2016) argues that as new cognitively potent technologies arise, it may be difficult to frame the ethical challenges we face from a deontological standpoint. It is, however, not always clear that this is the case. The guidelines developed by the High-Level Expert Group on Artificial Intelligence (AI HLEG)—an independent group established by the European Union—makes a notably deontological injunction into matters relating to the development and use of AI systems. The authors write:

> Human dignity encompasses the idea that every human being possesses an "intrinsic worth", which should never be diminished, compromised or repressed by others—nor by new technologies like AI systems. In this context, respect for human dignity entails that all people are treated with respect due to them as moral subjects, rather than merely as objects to be sifted, sorted, scored, herded, conditioned or manipulated. AI systems should hence be developed in a manner that respects, serves and protects humans' physical and mental integrity, personal and cultural sense of identity, and satisfaction of their essential needs. (AI HLEG 2019, p.10).

In the case of minds extended by AI systems (see Section 5), an adherence to this position would entail that AI systems should always work to the benefit of the extended subject, especially as regards their cognitive integrity and sense of identity.

---

9 See Clowes (2020), for several examples of how this might happen.

A commitment to individual and (especially) mental liberty, can also frame the ways technologies are used and appropriated. So, on the face of it, the charge that deontological approaches are relatively unresponsive to the challenges thrown up by technological innovation, as per Vallor, looks to be problematic. This is not to say, however, that deontology or consequentialism can offer completely satisfying vantage points.

A third approach to the consideration of ethical issues comes in the form of virtue ethics. Virtue ethics takes an alternative route to framing ethical concerns. In particular, it focuses on the notion of human flourishing—the ways in which humans can flourish and the conditions under which they can do so. Virtue ethics is distinct from both deontological and consequentialist approaches, in that it does not view ethics as primarily concerned with how to make the right decisions when confronted with ethical dilemmas. Instead, virtue ethics focuses on the development of the moral character of the individual and how he or she should live a good life. Decisions on character development are emphasized over decisions on individual actions.[10] It seeks to support the development of a person's moral character, so as to ensure that good decisions flow from their everyday practices and habits.

Virtue ethics builds upon several rich traditions in practical ethics to articulate values in which human flourishing can take place. In this context, it enjoins us to ask: How should we act in order to develop human excellences in the context of our new cognitive ecology? This might mean, for example, that we only come to rely on—or cognitively incorporate—technologies that afford a degree of reflective transparency. Reflective transparency, as we have noted, is a property afforded by at least some digital technologies. These are technologies where it is possible, at least in principle, for the user to understand the forces and factors that influence a technology's modus operandi. Such insights may help to attenuate some of the concerns raised in respect of mental manipulation and the biasing of an individual's thought and action. At the very least, understanding how a technology works, the situations in which it might fail, and the specific vulnerabilities that its further use might entail, serves as the informational bedrock that informs deliberate decisions about whether a particular form of bio-technological bonding is worth pursuing. Possessing such insights, an individual may decide that a given technology is not a suitable candidate for cognitive incorporation, or they may shift their focus to an alternative technology, one whose technological policies and practices are more closely attuned to the individual's ethical interests and concerns.

From a deontological or a consequentialist standpoint, a normative approach to the extended mind is apt to focus on the prevention of harm, especially as these harms relate to matters of personhood, autonomy, or individual agency. These vantage points might be of particular use when it comes to the decisions undertaken by policymakers or technology vendors. By contrast, the ability to develop a human-centred approach to the creation and sustenance of value-laden actions and habits may make virtue ethics a more suitable viewpoint for individuals seeking to regulate their own relations with cognitive technology. Virtue ethics can help us articulate ways for individuals to develop their own cognitive (and perhaps moral) characters in situations where technology is not merely external to us but also poised to become a part of us. It might help us think through what we want from technologies that have the potential to enhance or diminish our characters. And it might help inculcate a set of

10 Thanks to Jan-Hendrik Heinrichs for pointing this out.

practices and habits that enable us to press maximal cognitive and epistemic benefit from technologies, while, at the same time preserving our individual agency and autonomy.

Earlier in this chapter we discussed the importance of human agency and the conditions under which it might be compromised because of cognitive extension. Such considerations raise important questions about what sorts of cognitive beings we are, and, perhaps more importantly, the sorts of cognitive beings we want to be. Virtue ethics is a normative framework that is arguably well-suited to articulating many of the difficulties we face when we consider the nature of emerging digital technologies, many of which have their own autonomy and agency, and few of which are entirely transparent as regards their inner workings. When such technologies are poised to act as the *bona fide* constituents of our own mental machinery, then the stances we adopt towards those technologies become increasingly crucial. Just as it helps to have a degree of informed discretion about what food one consumes, a capacity to judiciously evaluate the potential ingredients of our future mental/cognitive architecture may be similarly important. Relative to the consequentialist and deontological frameworks, virtue ethics may have many advantages when it comes to considering such issues.

Claims about the extended mind alter the way we think about technological and artefactual resources and our relationships with them, introducing us to an unfamiliar and complex new ethical landscape. In the wake of such complexity, some may be inclined to renege on a commitment to the extended mind, opting instead for the seemingly less contentious view adopted by the proponents of embedded cognition. By assuming that artefacts are never, in fact, constituents of an individual's mind, complex questions of responsibility, privacy, manipulation, and autonomy can appear to be avoided.

But this comforting picture is false for several reasons. Some ethical concerns survive the transition from extended to embedded (and vice versa). In respect of privacy, for example, we still confront a range of important issues regarding access to online content, even if such content should be seen to exert a merely causal influence on the thoughts and actions of certain individuals. And there remain significant concerns over privacy, even if the causal/constitutive distinction is left aside. More seriously, from the extended point of view, the embedded perspective risks ignoring or failing to do justice to central ethical concerns with regard to the autonomy and dignity of individual persons. From the embedded point of view, the extended perspective risks overinflating ethical claims about property and artefacts to claims about those concerning the autonomy of persons. One of the motivations for framing ethical issues around the idea of the extended mind is precisely to take account of new agential formations. If Otto makes more sense as a coherent agent and as a person when his notebook is considered a proper part of himself, then shouldn't we want to protect his privacy, autonomy, and agency by affording his extended resources similar protections? Embedded mind theorists will beg to differ arguing that significant legal protections are already enshrined in law regarding the property of individuals. Yet, as the human race comes to rely on an ever-expanding panoply of cognitive technologies many more of us may soon be inhabiting the sort of cognitive space occupied by Otto. As more of us become apparently "extended agents" through our new engagement with AI, the Web, and other digital technologies, many will likely feel that their "extended" resources should be afforded the same protections and ethical entitlements as those they believe ought to be granted to Otto

and his notebook? For those wedded to the embedded view of mind however such claims are unlikely to prove truly persuasive, but it will be interesting to see how increasingly apparently hybrid minds involving personalized AI will be accounted for and understood.

This paper has explored the debate over the ethical status of the uses of contemporary technologies in the context of the extended mind debate. If we have not been able to finally settle the debate, at least we have offered some new and increasingly pressing dilemmas for the theorist and an expanded toolkit to explore its contours a little more adequately as we move forward.

Acknowledgements

Parts of this work were presented at the workshop on Cognitive Ecologies in Lisbon July 2022 and in the workshop on Epistemic and Ethical Issues in the Mind-Technology Problem in February 2023. The authors would like to thank those present at these meetings and also Georg Theiner, Jan-Hendrik Heinrichs and an anonymous reviewer for their perceptive and useful comments on an earlier draft manuscript.

References

Alexander, L., & Moore, M. (2021) Deontological Ethics. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Winter 2021 ed.). Stanford University, Stanford, California, USA.

Andrada, G., Clowes, R. W., & Smart, P. R. (2022) Varieties of transparency: Exploring agency within AI systems. AI & Society.

Becker, R., Chokoshvili, D., Comandé, G., Dove, E. S., Hall, A., Mitchell, C., Molnár-Gábor, F., Nicolàs, P., Tervo, S., & Thorogood, A. (2023) Secondary Use of Personal Health Data: When Is It "Further Processing" Under the GDPR, and What Are the Implications for Data Controllers? European Journal of Health Law, 30, 129–157.

Blitz, M. J. (2010) Freedom of thought for the extended mind: Cognitive enhancement and the constitution. Wisconsin Law Review, 2010(4), 1049–1118.

Boniface, M., Carmichael, L., Hall, W., Pickering, B., Stalla-Bourdillon, S., & Taylor, S. (2020) A Blueprint for a Social Data Foundation. Web Science Institute, University of Southampton, Southampton, UK. (Ref: WSI White Paper #4)

Bratman, M. E. (2000) Reflection, planning, and temporally extended agency. The Philosophical Review, 109(1), 35–61.

Carter, J. A. (2021) Varieties of (Extended) Thought Manipulation. In M. J. Blitz & J. C. Bublitz (Eds.), The Law and Ethics of Freedom of Thought, Volume 1 (pp. 291–309). Springer, Cham, Switzerland.

Carter, J. A., Clark, A., & Palermos, S. O. (2018) New Humans? Ethics, Trust and the Extended Mind. In A. J. Carter, A. Clark, J. Kallestrup, O. S. Palermos & D. Pritchard (Eds.), Extended Epistemology (pp. 331–351). Oxford University Press, Oxford, UK.

Carter, J. A., & Palermos, S. O. (2016) The ethics of extended cognition: Is having your computer compromised a personal assault? Journal of the American Philosophical Association, 2(4), 542–560.

Clark, A. (2007) Soft Selves and Ecological Control. In D. Ross, D. Spurrett, H. Kincaid & G. L. Stephens (Eds.), Distributed Cognition and the Will: Individual Volition and Social Context (pp. 101–122). MIT Press, Cambridge, Massachusetts, USA.

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

Clark, A., & Chalmers, D. (1998) The Extended Mind. Analysis, 58(1), 7–19.

Clowes, R. (2015) Thinking in the Cloud: The Cognitive Incorporation of Cloud-Based Technology. Philosophy & Technology, 28(2), 261–296.

Clowes, R. W. (2020) The Internet Extended Person: Exoself Or Doppelganger? LÍMITE Interdisciplinary Journal of Philosophy & Psychology, 15(Article 22), 1–23.

Delacroix, S., & Veale, M. (2020) Smart Technologies and Our Sense of Self: Going Beyond Epistemic Counter-Profiling. In M. Hildebrandt & K. O'Hara (Eds.), Life and the Law in the Era of Data-Driven Agency (pp. 80–99). Edward Elgar Publishing, Northampton, Massachusetts, USA.

Donald, M. (1991) Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition. Harvard University Press, Cambridge, Massachusetts, USA.

Elliot, M., O'Hara, K., Raab, C., O'Keefe, C. M., Mackey, E., Dibben, C., Gowans, H., Purdam, K., & McCullagh, K. (2018) Functional anonymisation: Personal data and the data environment. Computer Law & Security Review, 34(2), 204–221.

Heersmink, R. (2015) Dimensions of integration in embedded and extended cognitive systems. Phenomenology and the Cognitive Sciences, 14(3), 577–598.

Heersmink, R. (2017a) Distributed cognition and distributed morality: Agency, artifacts and systems. Science and Engineering Ethics, 23(2), 431–448.

Heersmink, R. (2017b) Extended mind and cognitive enhancement: Moral aspects of cognitive artifacts. Phenomenology and the Cognitive Sciences, 16(1), 17–32.

Heinrichs, J.-H. (2017) Against strong ethical parity: situated cognition theses and transcranial brain stimulation. Frontiers in Human Neuroscience, 11(Article 171), 1–13.

Heinrichs, J.-H. (2021) Neuroethics, cognitive technologies and the extended mind perspective. Neuroethics, 14(1), 59–72.

High-Level Expert Group on Artificial Intelligence. (2019) Ethics Guidelines for Trustworthy AI. European Commission, Brussels, Belgium. (https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai)

Hutchins, E. (1995) Cognition in the Wild. MIT Press, Cambridge, Massachusetts, USA.

Kraska, T., Stonebraker, M., Brodie, M., Servan-Schreiber, S., & Weitzner, D. (2019) SchengenDB: A data protection database proposal. In V. Gadepally, T. Mattson, M.

Stonebraker, F. Wang, G. Luo, Y. Laing & A. Dubovitskaya (Eds.), Heterogeneous Data Management, Polystores, and Analytics for Healthcare (pp. 24–38). Springer, Cham, Switzerland.

Levy, N. (2007) Neuroethics: Challenges for the 21st Century. Cambridge University Press, Cambridge, UK.

Mansour, E., Sambra, A. V., Hawke, S., Zereba, M., Capadisli, S., Ghanem, A., Aboulnaga, A., & Berners-Lee, T. (2016) A demonstration of the solid platform for social web applications. Proceedings of the 25th International Conference Companion on World Wide Web, Montreal, Quebec, Canada.

Mill, J. S. (1861/1998) Utilitarianism. Oxford University Press, New York, New York, USA.

O'Hara, K. (2020) Data Trusts. European Data Protection Law Review, 6(4), 484–491.

Palermos, S. O. (2023) Data, Metadata, Mental Data? Privacy and the Extended Mind. AJOB Neuroscience, 14(2), 84–96.

Patel, R., Wee, S. N., Ramaswamy, R., Thadani, S., Tandi, J., Garg, R., Calvanese, N., Valko, M., Rush, A. J., Rentería, M. E., Sarkar, J., & Kollins, S. H. (2022) NeuroBlu, an electronic health record (EHR) trusted research environment (TRE) to support mental healthcare analytics with real-world data. BMJ Open, 12(4), e057227.

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

Rupert, R. (2004) Challenges to the Hypothesis of Extended Cognition. Journal of Philosophy, 101(8), 389–428.

Sinnott-Armstrong, W. (2021) Consequentialism. In E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy (Fall 2021 ed.). Stanford University, Stanford, California, USA.

Smart, P. R. (2021) Shedding Light on the Extended Mind: HoloLens, Holograms, and Internet-Extended Knowledge. Frontiers in Psychology, 12(Article 675184), 1–16.

Smart, P. R., Clowes, R. W., & Heersmink, R. (2017) Minds Online: The Interface between Web Science, Cognitive Science and the Philosophy of Mind. Foundations and Trends in Web Science, 6(1–2), 1–232.

Taylor, S., Surridge, M., & Pickering, B. (2021) Regulatory Compliance Modelling Using Risk Management Techniques. 2021 IEEE World AI IoT Congress (AIIoT), Seattle, Washington, USA.

Vallor, S. (2016) Technology and the Virtues: A Philosophical Guide to a Future Worth Wanting. Oxford University Press, Oxford, UK.

Van Kleek, M., & O'Hara, K. (2014) The Future of the Social is Personal: The Potential of the Personal Data Store. In D. Miorandi, V. Maltese, M. Rovatsos, A. Nijholt & J. Stewart (Eds.), Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society. Springer, Berlin, Germany.

Vold, K., & Hernández-Orallo, J. (2022) AI Extenders and the Ethics of Mental Health. In F. Jotterand & M. Ienca (Eds.), Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues (pp. 177–202). Springer, Cham, Switzerland.

Weitzner, D. J., Hendler, J., Berners-Lee, T., & Connolly, D. (2006) Creating a Policy-Aware Web: Discretionary, Rule-Based Access for the World Wide Web. In E. Ferrari & B. Thuraisingham (Eds.), Web and Information Security (pp. 1-31). IRM Press, Hershey, Pennsylvania, USA.

Wheeler, M. (2019) The reappearing tool: transparency, smart technology, and the extended mind. AI & Society, 34(4), 857–866.