Erkenntnis
https://doi.org/10.1007/s10670-020-00320-6

ORIGINAL RESEARCH

Why the Self Does Not Extend

Keith Raymond Harris[1]

Received: 5 March 2020 / Accepted: 25 August 2020

Abstract
The defensibility of the extended mind thesis (EMT) is often thought to hinge on the possibility of extended selves. I argue that the self cannot extend and consider the ramifications of this finding, especially for EMT. After an overview of EMT and the supposed cruciality of the extended self to the defensibility of the former thesis, I outline several lines of argument in support of the possibility of extended selves. Each line of argument appeals to a different account of diachronic personal identity. I argue that no such argument for extended selves succeeds, as no account of diachronic personal identity is both plausible and supports the view that the self can extend. Next, I consider three objections that, if successful, would undercut the preceding argument that the self cannot extend. I conclude by reflecting on the implications of the conclusion that the self cannot extend, including the prospects for EMT.

1 Introduction

The *extended mind thesis* (EMT) raises questions about personal identity and the boundaries of the self. Some proponents of EMT maintain that the self can extend and that the defensibility of EMT depends on the capacity of selves to extend. In what follows, I dispute the former claim and comment briefly on the implications of this conclusion, especially for EMT.

After an overview of EMT and the supposed cruciality of the extended self to the defensibility of the thesis, I outline several lines of argument in support of the extended self. Each line of argument appeals to a different account of diachronic personal identity. I argue that no such argument for the extended self succeeds, as no account of diachronic personal identity is both plausible and supports the view that the self can extend. Next, I consider three objections that, if successful, would undercut the preceding argument that the self cannot extend. I conclude by reflecting on the prospects for EMT in light of the conclusion that the self cannot extend.

✉ Keith Raymond Harris
keithraymondharris@gmail.com

[1] University of Iowa, 570 English-Philosophy Building, Iowa City, USA

Published online: 12 September 2020

2 The Extended Mind and the Extended Self

According to EMT, mental states can sometimes be partially realized in features of the environment.[1] In what follows, I sometimes abbreviate this position by saying that the mind extends into the environment. Because the argument for EMT is familiar territory, I rehearse the argument only briefly here. The argument invokes a functionalist approach to the mental on which there is no in-principle barrier to mental properties being (partially) realized in features of the environment. This medium-neutral approach to the mental is captured by Clark and Chalmers's (1998) *parity principle*:

> *Parity principle*
> If, as we confront some task, a part of the world functions as a process which, *were it done in the head*, we would have no hesitation in recognizing as part of the cognitive process, then that part of the world *is* (so we claim) part of the cognitive process.

The *parity principle* is best understood not as providing a direct argument for EMT, but as a means of confronting and discouraging the common but unfounded conviction that the mental is confined to the body (Clark 2008, pp. 77–78).

The core of the argument for EMT is that features of the environment can play, for individuals, a functional role relevantly like the role of ordinary, brainbound beliefs.[2] This occurs, roughly, if the feature of the environment is readily accessible, used consistently, and not subject to scrutiny by the subject (Clark and Chalmers 1998, p. 17). These characteristics—sometimes labeled *availability*, *constancy*, and *automatic endorsement*, respectively—are not to be treated as necessary and sufficient conditions on an individual's mental states being partially realized in features of his environment, but provide an approximation of when extended beliefs are present. Clark and Chalmers illustrate the satisfaction of these conditions with the case of Otto who, suffering from early-stage Alzheimer's disease, stores information in a notebook and retrieves it as needed (1998). According to Clark and Chalmers, Otto's beliefs are partially realized in the notebook.

Having briefly reviewed EMT, we may turn to the significance of the extended self to the defense of EMT. Clark and Chalmers write the following about the extended self in their central illustrative case:

[1] Critics of active externalism often direct their attention toward the stronger thesis that mental states sometimes *are* realized in features of the environment (Rupert 2004, 2009). I adopt a weaker interpretation of EMT here and a correspondingly weak interpretation of the *extended self thesis* (EST) below. In challenging the weak interpretation of EST, I will thereby challenge stronger versions of that thesis and of EMT.

[2] It is a conspicuous feature of both critical and supportive discussions of EMT that these discussions typically focus on extended memory and extended belief, rather than extended mental states more generally—but see Susan Hurley (1998, 2010), Clark (2008), and Carter et al. (2016) for some notable exceptions. For present purposes, we need not go beyond this narrow focus.

> The information in Otto's notebook, for example, is a central part of his identity as a cognitive agent…Otto *himself* is best regarded as an extended system, a coupling of biological organism and external resources. To consistently resist this conclusion, we would have to shrink the self into a mere bundle of occurrent states, severely threatening its deep psychological continuity. (1998, p. 18 *original emphasis*)

Elsewhere, Clark is more reticent about extended selfhood:

> Does the putative spread of mental and cognitive processes out into the world imply some correlative (and surely unsettling) leakage of the self into the local surroundings? The answer now looks to be (sorry!) "Yes and No." No, because (as has already been conceded) conscious contents supervene on individual brains. But Yes, because such conscious episodes are at best snapshots of the self considered as an evolving psychological profile. (1997, p. 216),

In contrast to Clark's (1997) hesitance, Clark and Chalmers here express clear support for the proposition that, not only does Otto's mind extend into the notebook, so too does his self. Clark's initial ambivalence about extended selves is explicitly due to an uncertainty as to whether there is more to the self than one's conscious states—an uncertainty not shared by Clark and Chalmers. These brief remarks on the extended self allow room for various interpretations of what I will call the *extended self thesis* (EST).

To get a grasp on this thesis, it will be helpful to begin with some brief comments on the self. Philosophical debates broadly centered on the self are wide-ranging, but the two questions most relevant to the present discussion are the following:

(1) What are human persons?[3]
(2) What conditions must be satisfied for two human persons, existing at different times, to be the same person?

Philosophers interested in the self entertain a wide range of options to question (1). The issue is of interest here because some live answers to (1) quickly rule out the possibility of extended human selves. For instance, if human persons are organisms (Olson 2007) or are brains or brain hemispheres, then no human self extends. This is because neither human brains nor human organisms ever include non-biological features of the environment as parts.

Lively philosophical debate persists among champions of various answers to (1). Rather than providing an exhaustive overview of the state of the debate here, I will briefly discuss two live options for answering (1) that leave open the in-principle possibility of extended selves. The first such account of the self is the one apparently favored by Clark (1997) and Clark and Chalmers (1998).[4] On this account, the self

[3] For a thorough discussion of possible answers to question (1), see Olson (2007).

[4] In the passage quoted above, Clark and Chalmers actually suggest two distinct accounts of the self. The view that Otto is a coupling of an organism and external resources is inconsistent with the view that Otto is a bundle of mental states. To reconcile the authors' apparently commitment to two distinct accounts, I take Clark and Chalmers apparent commitment to the first account of the self to be shorthand for the view that Otto is realized or constituted by an organism together with external resources. Such short-

is an evolving bundle of mental states. Clark's (1997) initial ambivalence to EST is apparently due to some uncertainty as to whether the relevant bundle includes non-conscious states. But, as Clark and Chalmers (1998) suggest, where the relevant bundle is taken to include non-conscious states, there is a case to be made for EST. The suggestion here seems to be that, if the self is a bundle of conscious and non-conscious mental states, and some of these are realized beyond biological boundaries, then the self is partially realized beyond biological boundaries.

There is a second approach to human persons that leaves open the possibility of extended human selves. To introduce this approach, consider first Lynne Rudder Baker's account of the human person according to which such persons are constituted by, but not identical to, human organisms (2000). An account on which human persons are invariably constituted by human organisms is no more conducive to EST than an account on which human persons are human organisms or human brains. However, we can imagine a close variant on Baker's approach according to which human persons are often, but not invariably, constituted by human organisms. Such an account leaves open the possibility of human persons being partially constituted by features of the environment beyond the boundaries of the human organism, and so does not rule out EST.

Having recounted two live answers to (1) that do not by themselves rule out the possibility of extended selves, we can now state EST more clearly:

> *Extended Self Thesis*
> Human persons can be realized or constituted by human organisms together with features of their environments.

In many contexts, it is important to distinguish clearly between the relations of realization and constitution. However, in order to give EST the best possible chances of success, I formulate EST to be noncommittal about the relation between extended human persons and lower-level physical systems.

At this point, I want to make two terminological points that will simplify the discussion to follow. First, for the sake of simplicity and because it best fits Clark and Chalmers's (1998) approach to the self, I will primarily use the language of realization to capture the relation between human persons and the physical systems underlying them. This use of language should not be taken as indicating preference for the realization view over the constitution view. Second, I will use the terms 'human person' and 'biological person' to refer to individuals fully realized by organisms and the terms 'hybrid individual' and 'hybrid subject' to refer to individuals realized by organisms together with features of their environments.

As we will see below, some philosophers have attempted to argue for EST by way of appealing to certain answers to question (2). Roughly, their strategy has been to argue that the relation of personal identity sometimes holds between human persons

Footnote 4 (continued)
hands are widely employed and typically harmless in the context of discussions of EMT, but it will be necessary to avoid them here.

and hybrid individuals. Before considering this line of argument, let us look more closely at how EST has been understood by defenders and critics of the thesis.

Clark and Chalmers employ the language of parthood to capture the relation between Otto and the information in his notebook. More recently, in their critiques of EST, Baker (2009) and Olson (2011) employ similar language to capture the supposed relation between Otto and the notebook.[5] There is some cause for hesitation about the language of parthood, at least in the latter case.[6] Parthood is generally construed as a transitive relation such that, for any objects, A, B, and C, if A is a part of B and B is a part of C, then A is a part of C. When we attend to the transitivity of parthood, the parthood interpretation of the EST has some counterintuitive consequences. If we suppose that Otto's notebook is a part of Otto, and the front cover of the notebook is a part of the notebook, it follows that the front cover is a part of Otto. A more general concern is that, strictly speaking, the parthood interpretation of EST involves a category mistake. As we have seen, EST is most charitably understood as the thesis that human persons are sometimes realized or constituted by physical systems that include human organisms and features of their environments. On this understanding of EST, features of the environment are never parts of human persons, but are at most parts of the physical systems that realize or constitute human persons.

Having issued this caveat, I will sometimes use the terminology of parthood on the understanding that some object, A, may count as part of an object, B, in the relevant sense in virtue of being part of the realization of B, even where B is distinct from its realization. I will not assume that parthood in this loose sense obeys transitivity or the other axioms of classical mereology. This loose use of the language of parthood will suffice for present purposes as the substantive issues to be raised about EST will be independent of difficulties with applying the language of parthood to EST. With these caveats made, EST can loosely be stated as the thesis that the self can extend beyond its former boundaries such that features of the environment become parts of the self.

We now turn to the significance of EST to EMT. First, EST is arguably necessary to maintain functional parity between ordinary mental states and extended mental states. Consider, for example, the objection that Otto's use of the notebook is functionally different to ordinary cognizers' reliance on memory insofar as the former, but not the latter, involves perception of external objects (Butler 1998, p. 211). Clark and Chalmers (1998, p. 16) and later Clark (2010) respond that this objection misconstrues the Otto case. Here is how Clark describes the case:

[5] Whereas Baker (2009) and Olson (2011) dispute EST principally on the basis of its implausible implications, the thrust of the critique to follow is that EST is not well motivated. Because the critique to follow is largely independent of the objections raised by Baker and Olson, I do not comment on their objections at length here.

[6] Clark and Chalmers's suggestion that the *information* in Otto's notebook is, on the face of it, not subject to the sort of objection to which the suggestion that the notebook itself is part of Otto is subject.

> [F]rom our point of view, Otto's inner processes and the notebook constitute a single, extended cognitive system. Relative to *this* system, the flow of information is wholly internal and functionally akin to introspection. (2010, p. 57)

According to this response, the supposed functional difference between ordinary and extended mental states is illusory on the grounds that the notebook is a part of Otto. The resilience of EMT against the present line of objection thus depends on Otto himself being extended.

The second reason for which EST is sometimes taken to be essential to the defensibility of EMT concerns the functionalist justification for EMT. Kengo Miyazono (2017) alleges that functionalism fails to support the attribution of beliefs partially realized in the notebook to Otto and, more generally, fails to support EMT. Miyazono appeals to a popular functionalist response to Searle's (1980) *Chinese Room Argument* to motivate the objection. On the *systems reply* to Searle's argument, it is not the man within the room, but instead a system comprising the man and the instructions and rules that allow him to correlate the Chinese characters he receives with those he is to output that understands Chinese.[7] The thrust of the *systems reply* is that the subject of the understanding of Chinese is not some biological individual, but instead a system that has that individual and certain material objects in his environment as parts.[8] Miyazono then alleges that functionalists, at least those functionalists that endorse the *systems reply*, ought to conclude that whatever beliefs are partially realized in Otto's notebook are not Otto's beliefs, but instead the beliefs of a hybrid individual realized by Otto and his notebook.[9] Notably, Miyazono takes his argument to support not only the conclusion that Otto's mental states not partially realized in the notebook, but that extension of mind is probably metaphysically impossible (2017, pp. 3539–3540).

Milojevic (2018) follows Miyazono in allowing that the beliefs partially realized in Otto's notebook are the beliefs of a hybrid individual. However, Milojevic argues that, because Otto's self is extended to have the notebook as a part, the hybrid individual that has the notebook as a part *is* Otto. Milojevic thereby defends EMT from Miyazono's *systems reply* objection by appeal to EST.

[7] For an early and clear endorsement of the *systems reply*, see Jack Copeland (1993, ch. 6).

[8] In line with what has been said above, it is perhaps better to say that the subject that understands Chinese is not the system, but a hybrid individual realized by this system. In fact, some responses to the *Chinese Room Argument* emphasize a distinction between the Chinese room system and the subject it realizes (Cole 1991).

[9] Participants in the debate over EMT typically do not observe the distinction between persons and the physical systems that realize them. Here and below, I import the distinction and apply it even where the authors discussed do not.

3 The Extended Self and Personal Identity

We have seen that the defensibility of EMT is related in two ways to EST. First, certain direct objections to EMT apply if and only if the self cannot extend. Second, at least for those functionalists that take mental states to be the properties of the entire systems that realizes them, as is the case for those that endorse the *systems reply*, the functionalist case for the claim that Otto is the subject of extended mental states depends on the further claim that Otto himself extends.

I now consider a series of arguments for EST, based on answers to question (2) above, that is, based on distinct accounts of diachronic personal identity. I focus on the question of whether any of these accounts support the notion that Otto has an extended self. The accounts considered below represent a fraction of the many accounts of diachronic personal identity that enjoy philosophical prominence. Consequently, the mere demonstration that none of these accounts provides support for Otto's extension would provide only a weak basis on which to conclude that the self cannot extend. However, we will uncover reason to think that no plausible account of diachronic personal identity supports the claim that Otto's self extends or EST more generally.

We begin with an account of personal identity that is widely attributed to Locke (1975). On this view, some person B at time t_2 is the same person as some person A at earlier time t_1 just in case B remembers some experiences A has at t_1. The *memory view* offers a promising foundation for EST, insofar as the hybrid individual realized by Otto and the notebook plausibly remembers the experiences of Otto before he began to use the notebook. However, the *memory view* is subject to the objection that it must deny the transitivity of personal identity in cases where persons A, B, and C are such that C remembers some experiences of B and B remembers some experiences of A, but C does not remember any experiences of A.

That the *memory view* affirms that A = B and that B = C, while denying that A = C is sufficient to discount the view as a plausible account of personal identity over time. However, the *memory view* faces a further difficulty worth discussing here. There are a range of cases in which multiple apparently distinct future individuals remember the experiences of some single earlier individual. One such case involves an individual whose cerebrum is divided into halves, each of which is surgically implanted into a different body.[10] This procedure could in principle yield two apparently distinct individuals both of which remember the experiences of the person from whom their halves of the brain were taken. In such a case, the *memory view* yields the conclusion that both products of the division are identical to the earlier individual whose brain was divided.[11] Yet it is hardly plausible that the two products of the division are the same individual. This last claim earns credence both from general intuition and from the *memory view* itself, insofar as neither product of the

[10] For an early discussion of such cases, see Wiggins (1967). Fission cases were later discussed at length by Parfit (1984) and have since become a staple of the personal identity literature.

[11] The memory view ascribed here to Locke is not alone in having this unwelcome implication. See Perry (1972) for an overview of some accounts with the same implication.

division will remember experiences of the other product that occur after the division. Further cases that are problematic for the *memory view* can be generated by considering rather more outlandish causes of division. Parfit's (1984) case of the teletransporter, for instance, can be constructed in such a way that multiple apparently distinct individuals remember the experiences of some earlier individual.

The evident difficulties with the *memory view* motivate the conclusion that any plausible account of personal identity must respect the constraint that no individuals existing at a time can be distinct while being identical to the same individual existing at an earlier time. I turn now to a series of accounts of diachronic personal identity that obey this constraint. The accounts are representative of a broader distinction in accounts of diachronic personal identity between those that construe that relation in terms of some sort of psychological continuity and those that construe that relation in terms of physical continuity. Consider first the following criterion of personal identity (1984, p. 207):

> *The Psychological Criterion*
> (1) There is *psychological continuity* if and only if there are overlapping chains of strong connectedness. X today is one and the same person as Y at some past time if and only if (2) X is psychologically continuous with Y, (3) this continuity has the right kind of cause, and (4) it has not taken a 'branching' form. (5) Personal identity over time just consists in the holding of facts like (2) to (4).

The *psychological criterion* improves upon the memory view in two crucial respects. First, because psychological continuity can hold between individuals even when there is no *direct* psychological connection between these individuals, the *psychological criterion* does not face the difficulties with transitivity that are commonly raised in objection to the memory view. Second, the *psychological criterion* includes a non-branching condition that rules out personal identity over time in cases where multiple individuals are psychologically continuous with some earlier individual.

Because of these improvements over the memory view, the *psychological criterion* is, at least in relative terms, a plausible account of diachronic personal identity. Milojevic (2018) appeals to the *psychological criterion* to defend EST, arguing that the hybrid individual that includes the notebook as a part is psychologically continuous with Otto before he began to use the notebook. To assess Milojevic's line of argument, it will be helpful to establish a timeline and some abbreviations. Suppose t_1 is a time prior to when Otto first began to use the notebook, t_2 is the time at which Otto begins to use the notebook, and t_3 is some later time. Call the figure that begins to use the notebook at t_2 B-Otto, for biological Otto. Finally, call the hybrid individual realized by the organism and the notebook at t_3 N-Otto.

The difficulty with Milojevic's proposal is that, while N-Otto at t_3 is psychologically continuous with B-Otto prior to t_2, it seems that N-Otto is not alone in this respect (Harris 2019; Raymond Harris 2020). Even after t_2, it is plausible that there is some human person who suffers from a deficiency in memory and who accesses the contents of the notebook by way of perception. Supposing B-Otto persists after beginning to use the notebook, there are two distinct entities that are plausibly

psychologically continuous with B-Otto prior to t_2, B-Otto and N-Otto.[12] Because N-Otto at t_3 is not uniquely psychologically continuous with B-Otto prior to t_2, it follows from application of the *psychological criterion* that N-Otto at t_3 is not identical to B-Otto at t_1.[13]

That the *psychological criterion* fails to support the extension of Otto is not an idiosyncratic consequence that can be avoided by other plausible theories of diachronic personal identity. Consider next the *narrative view* of personal identity. Like the *psychological criterion*, the *narrative view* is an offshoot of the Lockean *memory view*. On the *narrative view*, what matters to personal identity over time is not access to the past by way of some sterile repository of episodic memories. Rather, what matters is the location of oneself in a coherent first-personal narrative condensed from recollections of events. Some versions of the *narrative view* depart considerably from rival offshoots of the Lockean view (Schechtman 1994, 1996), while others are best understood as refinements of that view (Schroer and Schroer 2014). However, any plausible version of the *narrative view* must grapple with questions of fission (Reid 1997, pp. 216–218). More specifically, any plausible version of the *narrative view* must include a non-branching condition. Schroer and Schroer (2014) defend the following version of the *narrative view,* which obeys this constraint:

> *The Narrative View*
> An earlier person stage X and a later person stage Y are two stages of the same person iff: (1) There is narrative continuity (which is composed of narrative connectedness or overlapping chains of narrative connectedness) between some of the mental states/actions of X and some of the mental states/actions of Y, (2) These mental states are causally related to each other in the right way, and, (3) There is no branching. (2014, p. 463)

The inclusion of (3) in Schroer and Schroer's version of the *narrative view* is motivated by just the sort of fission cases that motivated Parfit's non-branching condition (2014, p. 467).

Heersmink (2017, 2018) has recently argued from the *narrative view* to EST. Heersmink develops this argument, in part, by appealing to emerging lifelogging technologies. These technologies, which may record information in the form of images, audio, video, or text, are used both by patients with cognitive impairments, including patients with Alzheimer's disease (2017, pp. 3145–3146), and by individuals simply aiming to generate a more complete record of their lives than biological memory alone would allow (2017, p. 3147). Heersmink's discussion of lifelogging technologies provides a welcome update of extended memory technologies, as compared to the case of Otto and his notebook, but do such lifelogging technologies genuinely extend the self? According to Heersmink, the

[12] Given Otto's condition, it may well be that N-Otto at t_3 is, relative to B-Otto at t_3, connected to B-Otto at t_1 by denser chains of psychological connectedness. Even so, it remains the case that both later entities are psychologically continuous with B-Otto at t_1.

[13] Here we should acknowledge the consequence that, on the *psychological criterion*, B-Otto at t_3 is also not quantitatively identical to B-Otto at t_1 (Harris 2019).

memories that constitute individual's personal narrative are not only prompted by, but may be realized in, external artifacts including lifelogging technologies. As a consequence, Heersmink maintains that the self is partially constituted by such artifacts (2018, p. 1830).

To assess Heersmink's argument from the *narrative view* to EST, we need not assess the general merits of the *narrative view* here. Independent of its overall plausibility, the *narrative view* fails to support EST. Consider the following revealing passage from Heersmink:

> A critic might argue that even if we are isolated from all our objects and other people such that there is no material and social scaffolding of memory, *there remains a core self. This self may be rather diminished and may have a much less detailed and stable narrative, but still has a narrative.* I agree with this view. (2018, p. 1845 *emphasis added*)

A hybrid individual realized by some biological individual and her lifelogging technologies may be the subject of an unfolding narrative, in the sense that the memories constitutive of this narrative may be distributed across biological and non-biological parts of the underlying system. Indeed, this narrative may link together the hybrid individual to the purely biological person that preceded it. But, crucially, even when a biological individual immerses herself in an environment that partially realizes the narrative of the hybrid subject, the biological individual, the core self, remains. Consequently, given the non-branching condition, those hybrid subjects that are narratively continuous with biological individuals are nonetheless distinct individuals. To illustrate, let us return to Otto, and suppose that the notebook serves a similar purpose to the one served by the lifelogging technologies described above. We may say that a narrative linking N-Otto at t_3 to B-Otto at t_1 fails to secure the identity of these person stages precisely because there is a further individual at t_3, distinct from N-Otto, who is also linked by narrative continuity to B-Otto at t_1. In short, the *narrative view* fails to support EST for almost precisely the same reason that the *psychological criterion* failed. This result is not altogether surprising given that, as we have seen, some proponents of the *narrative view* take this to be a refinement, rather than a replacement, of existing neo-Lockean views.

Consider a final account of personal identity that differs radically from those considered thus far (Parfit 1984, p. 204):

> *The Physical Criterion*
> (1) What is necessary is not the continued existence of the whole body, but the continued existence of *enough* of the brain to be the brain of a living person. X today is one and the same person as Y at some past time if and only if (2) enough of Y's brain continued to exist, and is now X's brain, and (3) this physical continuity has not taken a 'branching' form. (4) Persona1 identity over time just consists in the holding of facts like (2) and (3).

Plausibly enough, Otto's brain persists between t_1 and t_3 and is N-Otto's brain at t_3. There is physical continuity between B-Otto prior to t_2 and N-Otto at t_3.

However, the *physical criterion*, like its psychological counterpart, includes a non-branching condition. The inclusion of this condition is not incidental. As we have seen in the course of the discussion of the memory view, there can be no plausible account of personal identity over time that lacks a non-branching condition. Yet, supposing that B-Otto exists at t_3, no account of diachronic personal identity that includes a non-branching condition has the implication that Otto extends. Because any plausible account of personal identity over time must include a non-branching condition, no plausible account of personal identity can support the claim that Otto extends. More generally, any attempt to defend EST by appeal to criteria of diachronic personal identity faces a dilemma. A criterion of personal identity can include a non-branching condition to rule out counterexamples or it can support EST, but it cannot do both.

4 Objections and Replies

I have argued that no plausible account of personal identity has the implication that Otto extends. The reasoning adduced to support this conclusion is general and supports the broader conclusion that the self cannot extend. In this section, I consider three possible avenues for resisting this latter conclusion.

Notice first that the argument developed against EST above depends heavily on non-branching conditions in accounts of diachronic personal identity. I argued that Otto's self does not extend because, while N-Otto at t_3 is continuous, both psychologically and physically, with B-Otto prior to t_1, the former is not uniquely continuous with the latter in these ways. Here one might object that, while the non-branching condition applies in this case, the applicability of the non-branching condition is peculiar to the Otto case. The conclusion drawn in the Otto case does not generalize, because, following a supposed process of extension, there need not be a biological individual that is continuous with the individual whose mind is supposedly extended. If this is right, then there can be cases in which the non-branching condition fails to rule out the diachronic identity of some biological individual and some hybrid individual. Thus, the argument above fails to demonstrate that EST is false, or so the objection goes.

This line of objection fails, however, because the existence of a biological individual at t_3 in the Otto case that is continuous with the biological individual at t_1 is not a mere quirk of that case. Consider how Clark and Chalmers describe putative cases of extended cognition:

> In these cases, the human organism is linked with an external entity in a two-way interaction, creating a *coupled system* that can be seen as a cognitive system in its own right. (1998, p. 8)

The point is underscored by recent applications of *dynamic systems theory* (DST) to the study of extended cognitive systems (Palermos 2014). The tools of DST allow for a biological organism and a feature of its environment to be modeled as distinct systems whose interactions give rise to a coupled system with both entities as parts, and whose properties cannot be reduced to the properties of its parts. Crucially, the

existence of the coupled system depends on the existence of its component systems, one of which is a biological organism. Similarly, the existence of the hybrid individual constituted by the coupled system depends on the existence of the biological organism that fully realizes the human person. Thus, in contrast to the cases of division most common to the personal identity literature, the products of division in apparent cases of extended mentality are such that the existence of one product depends on the existence of the other. Branching, in these latter cases, is inevitable.

The case I have developed against EST thus far depends on the non-branching condition on diachronic personal identity, a condition that is itself motivated by the plausible claim that no two distinct individuals can be diachronically identical to the same earlier individual. The defender of EST might object, though, that this claim can be respected without commitment to a non-branching condition. The desired result can be achieved by some version of the *closest continuer view*, according to which the relation of diachronic personal identity solely between individuals and their unique closest continuers, if there are any (Nozick 1981). Various versions of the *closest continuer view* might be proposed, depending on what form of continuity one takes to be relevant. For the defender of EST, the most promising version of the view is probably one that emphasizes psychological or narrative continuity. After all, the closest physical continuer of B-Otto is undoubtedly B-Otto. The defender of EST might, for instance, argue that, owing to Otto's deficiencies in biological memory, the closest psychological and narrative continuer of B-Otto is N-Otto.[14]

There are significant difficulties with this strategy for defending EST, though. First, the strategy assumes a particular account of diachronic personal identity—one appealing to closest psychological or narrative continuity—that is not obviously well-motivated. I am not aware of any philosopher that explicitly endorses the sort of account required. While Nozick (1981), for instance, endorses a version of the *closest continuer view*, his version of the view does not treat any particular form of continuity, psychological or otherwise, as solely determinative of diachronic personal identity. Second, insofar as the strategy appeals to Otto's deficiencies in biological memory to motivate the view that a hybrid individual is the closest psychological or narrative continuer of a human person, the strategy can support the existence of extended selves only in a limited range of cases. It would not, for instance, support the view that human persons in general are 'natural born cyborgs' (Clark 2003). Finally, even in the Otto case, it is not clear that a hybrid individual is the closest psychological or narrative continuer of a biological person. It is important to recognize that, even in the Otto case, the successful use of the notebook requires that B-Otto remembers to carry the notebook, that the notebook is important for him, how to use it, and so on. Moreover, the deterioration of Otto's memory is itself an element of psychological and narrative continuity. To see this, notice that we have two candidates for diachronic personal identity with B-Otto following t_2. One of these candidates shares in B-Otto's deficiencies in memory. The other candidate has a radically expanded memory capacity. The disconnect between B-Otto prior to t_2 and N-Otto following t_2 is a reason to regard B-Otto following t_2 as the

[14] Thanks to an anonymous reviewer for pressing me on this point.

closest psychological and narrative continuer of B-Otto prior to that time, and thus is a reason to think that Otto himself follows the B-Otto branch. This consideration is not by itself decisive, for it may be that continuity in memory content is a more important element of psychological and narrative continuity for diachronic personal identity than continuity in memory capacity. However, this line of response shows, at least, that even a psychological or narrative version of the *closest continuer view*, applied to a peculiar case involving deterioration of memory, does not obviously support EST.

I want to conclude my response to the present objection with a concession and a challenge. First, what I have said here does not rule out the possibility that there is some version of the *closest continuer view* that supports the extension of the self in some small range of cases. For that matter, I have not entirely excluded the possibility that there is some further account of diachronic personal identity that supports the extension of the self in some cases. I have, however, presented reason to doubt that there is such an account, for I have argued that any plausible account of diachronic personal identity must exclude the possibility of two distinct individuals being identical to the same earlier individual and I have argued that the accounts that do this inevitably deny the identity of hybrid individuals with earlier human persons. Moreover, it is reasonable to think that the burden of demonstrating that there is some such account of diachronic personal identity, and that this account of diachronic personal identity is the correct one, lies with the defender of EST. After all, intuition and philosophical orthodoxy are strongly at odds with the claim that human persons can come to be partially realized by non-biological features of the human organism's environment. Any account of diachronic personal identity that entails this claim would thus need to be sufficiently well-motivated to avoid the conclusion that doing so reduces the account to absurdity.

Consider now one final maneuver that might be invoked to defend EST. Lewis (1976) proposes to defend the commonsensical claim that it matters that someone quantitatively identical to oneself exists in the future from Parfit's (1971) arguments to the contrary. For present purposes, Lewis's relevant claim is that in cases of division there are two distinct individuals who merely share person stages for a portion of their lives. Returning to the Otto case, one might argue that there are two distinct individuals Otto-1 and Otto-2 who coincide until t_2, after which Otto-1 remains a biological individual while Otto-2 becomes a hybrid individual composed of a biological individual together with non-biological features of his environment. On this view, Otto-1 does not extend but Otto-2 does. EST is saved, or so one might think.

There are at least two serious difficulties with an appeal to Lewis-style considerations in this context. First, while I will not rehearse the objections here, Lewis's response to Parfit is highly controversial (Parfit 1976; Roache 2010). Second, on the Lewisian reply, there are two individuals in the Otto case, one of which comes to have an extended mind and the other of which does not. But, on the functionalist framework that motivates EMT, there is no basis on which to think that one of the individuals in question has an extended mind while the other does not. Whatever functional relation Otto-2 comes to stand in with respect to the notebook, Otto-1 comes to stand in precisely the same functional relation to the notebook (cf. Olson 2011, p. 487). Consequently, the functionalist underpinning of EMT cannot support

the conclusion that just one of the entities that temporarily coincides with the other comes to be extended.

5 Conclusion: The Self Cannot Extend, Does That Matter?

Parfit (1984) memorably argued that personal identity is not what matters in survival. We have considered a series of unsuccessful arguments for EST, and have uncovered good reason to think that the self cannot extend. Supposing this conclusion is correct, does it matter? Defenders of EMT might think so because, as we have seen, it is sometimes thought that the defensibility of EMT requires EST. Yet the implications for EMT in light of the argument developed here are unclear, as it has recently been argued that the most defensible version of EMT does not require extended selves (Harris 2019).

The denial of EST might be thought to have further significance. Carter and Palermos (2016) have recently considered the ethical implications of EST. Supposing that selves may extend into one's environment, acts that might ordinarily be construed as interference with a person's tools are better understood as forms of personal assault. The argument developed here suggests that selves do not so extend, but does this matter? Perhaps not. First, to deny that selves extend is not to deny that there are hybrid subjects with biological individuals and non-biological components as parts. These hybrid subjects may well be the subjects of personal assault, even if their biological components are not. Second, as Parfit thought, identity may not be what matters in survival. Although we cannot become hybrid subjects, we may well have cause for intimate concern for the interests of such subjects, insofar as they are connected to us by relations of psychological continuity.[15]

References

Baker, L. R. (2000). *Persons and bodies: A constitution view*. New York: Oxford University Press.

Baker, L. R. (2009). Persons and the extended mind thesis. *Zygon: Journal of Religion and Science, 44*(3), 642–658.

Butler, K. (1998). *Internal affairs: A critique of externalism in the philosophy of mind*. Dordrecht: Kluwer.

Carter, J. A., Gordon, E. C., & Palermos, S. O. (2016). Extended emotion. *Philosophical Psychology, 29*(2), 198–217.

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

Clark, A. (1997). *Being there: Putting brain, body, and the world together again*. Cambridge: MIT Press.

Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. New York: Oxford University Press.

[15] For a concrete example, see Daniel Dennett (1996, p. 138) and Heersmink's (1845–1846) discussion of the removal of Alzheimer's patients from their homes. The foregoing discussion suggests that while such acts do not amount to a dismantling of the self—at least not the self of any biological individual with Alzheimer's disease—such acts may be equally as concerning as if they involved compromise of the self.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.
Clark, A. (2010). Memento's revenge: The extended mind extended. In R. Menary (Ed.), *The extended mind* (pp. 43–66). Cambridge: MIT Press.
Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*(1), 7–19.
Cole, D. (1991). Artificial intelligence and personal identity. *Synthese, 88*(3), 399–417.
Copeland, J. (1993). *Artificial intelligence: A philosophical introduction*. Malden, MA: Wiley-Blackwell.
Dennett, D. (1996). *Kinds of minds: Towards and understanding of consciousness*. New York: Basic Books.
Harris, K. (2019). Whose (extended) mind is it, anyway? *Erkenntnis*. https://doi.org/10.1007/s10670-019-00172-9.
Heersmink, R. (2017). Distributed selves: Personal identity and extended memory systems. *Synthese, 194*(8), 3135–3151.
Heersmink, R. (2018). The narrative self, distributed memory, and evocative objects. *Philosophical Studies, 175*(8), 1829–1849.
Hurley, S. (1998). *Consciousness in action*. Cambridge: Harvard University Press.
Hurley, S. (2010). Varieties of externalism. In R. Menary (Ed.), *The extended mind* (pp. 101–153). Cambridge: MIT Press.
Lewis, D. (1976). Survival and identity. In A. O. Rorty (Ed.), *The identities of persons* (pp. 17–40). Berkeley: University of California Press.
Locke, J. (1975). An essay concerning human understanding. In P. H. Nidditch (Ed.), *The clarendon edition of the works of John Locke*. Oxford: Oxford University Press.
Milojevic, M. (2018). Extended mind, functionalism, and personal identity. *Synthese*. https://doi.org/10.1007/s11229-018-1797-5.
Miyazono, K. (2017). Does functionalism entail extended mind? *Synthese, 194,* 3523–3541.
Nozick, R. (1981). *Philosophical explanations*. Cambridge: Harvard University Press.
Olson, E. T. (2007). *What are we? A study in personal ontology*. New York: Oxford University Press.
Olson, E. T. (2011). The extended self. *Minds and Machines, 21*(4), 481–495.
Palermos, S. O. (2014). Loops, constitution, and cognitive extension. *Cognitive Systems Research, 27*(1), 25–41.
Parfit, D. (1971). Personal identity. *The Philosophical Review, 80*(1), 3–27.
Parfit, D. (1976). Lewis, Perry, and what matters. In A. O. Rorty (Ed.), *The identities of persons* (pp. 91–107). Berkeley: University of California Press.
Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
Perry, J. (1972). Can the self divide? *The Journal of Philosophy, 69*(16), 463–488.
Raymond Harris, K. (2020). Group minds as extended minds. *Philosophical Explorations*, *23*(3), 234–250.
Reid, M. (1997). Narrative and fission: a review essay of Marya Schectman's *The constitution of selves*. *Philosophical Psychology, 10*(2), 211–219.
Roache, R. (2010). Fission, cohabitation, and the concern for future survival. *Analysis, 70*(2), 256–263.
Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy, 101*(8), 389–428.
Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.
Schechtman, M. (1994). The truth about memory. *Philosophical Psychology, 7*(1), 3–18.
Schechtman, M. (1996). *The constitution of selves*. Ithaca, NY: Cornell University Press.
Schroer, J. W., & Schroer, R. (2014). Getting the story right: a Reductionist narrative account of personal identity. *Philosophical Studies, 171,* 445–469.
Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–457.
Wiggins, D. (1967). *Identity and spatio-temporal continuity*. Oxford: Oxford University Press.