Varieties of the extended self

Richard Heersmink

Forthcoming in Consciousness & Cognition

Abstract
This article provides an overview and analysis of recent work on the extended self, demonstrating that the boundaries of selves are fluid, shifting across biological, artifactual, and sociocultural structures. First, it distinguishes the notions of minimal self, person, and narrative self. Second, it surveys how philosophers, psychologists, and cognitive scientists argue that embodiment, cognition, emotion, consciousness, and moral character traits can be extended and what that implies for the boundaries of selves. It also reviews and responds to various criticisms and counterarguments against the extended self. The main focus is on the link between the extended mind and extended self, which has received the most attention in recent literature. But accounts of the extended self developed independently of the extended mind are also briefly discussed. This article ends by drawing out some of the conceptual, methodological, and normative implications of the extended self and suggesting some directions for future research.

Keywords: Extended self; Distributed self; Soft self; Material self; Identity; Embodiment; Extended mind; Extended emotion; Extended character; Extended consciousness

1. Introduction
Philosopher and psychologist William James (1890), in his influential book The Principles of Psychology, develops the notion of an empirical self.

“The Empirical Self of each of us is all that he is tempted to call by the name of me. But it is clear that between what a man calls me and what he simply calls mine the line is difficult to draw. We feel and act about certain things that are ours very much as we feel and act about ourselves” (James 1890, p. 291).

James is interested in the boundaries of the empirical self and goes on to argue that the empirical self is partly constituted by artifacts and other people.

“In its widest possible sense, however, a man's Self is the sum of all that he can call his, not only his body and his psychic powers, but his clothes and his house, his wife and children, his ancestors and friends, his reputation and works, his lands, and yacht and bank account. All these things give him the same emotions. If they wax and prosper, he feels triumphant; if they dwindle and die away, he feels cast down, not necessarily in the same degree for each thing, but in much the same way for all [...] An instinctive impulse drives us to collect property; and the collections thus made become, with different degrees of intimacy, parts of our empirical selves” (James 1890, p. 291-93).
James here argues that artifacts and other people can become part of the empirical self, because they generate affective states in the agent, in that way suggesting a constitutive relation between self, emotion, and environment. So, on James’ view, external artifacts and other people are co-constitutive of our empirical self, because they cause emotions1.

The notion that the self extends into environmental resources or is distributed across embodied agents and environmental resources has recently been developed in a number of novel ways. This article provides an overview and analysis of the recent literature on the extended and distributed self, with a particular focus on the link between the extended mind and the extended self. Some theorists argue that the extended mind implies an extended self (Clark and Chalmers 1998; Clark 2003, 2007a, b; Anderson 2008; Malafouris 2008; Hongladarom 2016; Howell 2018; Piredda & Candiotto 2019; Milojevic 2020; Clowes forthcoming), whereas others deny this (Wilson 2004; Baker 2009; Olsen 2011; Wilson & Lenart 2014; Harris 2019). Other theorists, however, develop a notion of an extended self independent of the extended mind (James 1890; Belk 1988; Wallace 2019a, 2019b).

Either way, various theorists and philosophers have pointed out that the self is not a merely neurological and biological phenomenon, but should be conceptualised as a relational, extended, and distributed entity. This ontological claim has methodological and normative implications. To examine this entity, we should enlarge the unit of analysis in the conceptual and empirical study of the self, focusing on agent-environment interactions. And if environmental resources are genuinely part of the self, they obtain the same moral status as neurological and biological parts of the self, suggesting we ought not to interfere with and respect people’s extended selves.

In what follows, I start by briefly characterising the target domain, outlining the notions of minimal self, person, and narrative self (Section 2). I then respectively present how some theorists argue that embodiment (Section 3), cognition (Section 4), emotion (Section 5), consciousness (Section 6), and moral character traits (Section 7) can be extended and what that implies for the boundaries of selves. I end with a concluding reflection, drawing out some of the conceptual, methodological, and normative implications of the extended self and suggesting some directions for future research (Section 8).

2. Minimal selves, persons, and narrative selves

Concepts of self have a rich and varied history in philosophy, psychology, psychiatry, neuroscience, and related fields (e.g., Barresi & Martin 2011; Gallagher 2011). Self has been

---

1 This may, however, not be sufficient to argue for an extended or distributed self. Several theorists have recently conceptualised emotion as extended (Slaby 2014; Colombetti & Roberts 2015; Carter, Gordon & Palermos 2016; Krueger & Szanto 2016). A central feature of genuinely extended emotions is that an external resource needs to be reciprocally integrated into emotional states and processes. A mere causal influence on emotional states and processes (ala James) will only show that emotion, and by extension self, are embedded and situated. So, to show that self is extended because emotion is extended, James needed to show that objects and people are reciprocally integrated into emotional states and processes. Giulia Piredda & Laura Candiotto (2019) recently used a Jamesian view to argue for an affectively extended self. Focusing on the nature of experience, they argue that reciprocal embodied interactions with certain personal objects extends the self, because extended emotions are constitutive of the self.
conceptualised as an illusion, subject of experience, stream of consciousness, centre of narrative gravity, autobiographical narrative, and as embodied, extended, dialogical, social, and historical. Different scholars thus talk about different things when they talk about the self. Eric Olson (1998) goes so far as to say there is no problem of “the self” as scholars cannot agree on the characteristic features of selves. There are ongoing debates in the humanities, social sciences, and neurosciences about how to characterise the self and how to conceptualise the relation between self, body, and the environment in which it is embedded. It’s not the goal of this article to provide an overview of this more general literature, but in order to better understand the current debate on the extended self, it is helpful to highlight some of these notions. I’ll look in particular at minimal/embodied views of self, personhood, and narrative/autobiographical views (see Gallagher 2000 and Zahavi 2010 for the distinction between the minimal and narrative self).

The minimal self is characterised by Shaun Gallagher as follows:

“Phenomenologically, that is, in terms of how one experiences it, a consciousness of oneself as an immediate subject of experience, unextended in time. The minimal self almost certainly depends on brain processes and an ecologically embedded body, but one does not have to know or be aware of this to have an experience that still counts as a self-experience”. It is the self that remains when “all of the unessential features of self are stripped away” and “is limited to that which is accessible to immediate self-consciousness (2000, p. 15).

Regarding the minimal self, Gallagher identifies two modalities of experience: (1) a sense of ownership of experience and (2) a sense of agency which is the sense that I am the author of the action (see de Haan & de Bruin 2009 for discussion). The minimal self thus has subjecthood and agenthood. In normal action, these coincide and are indistinguishable. They can, however, come apart, for example when someone else moves my arm, I have a sense of ownership of experience but not of agency. The sense of ownership of experience can be explained in terms of proprioception and perception. The minimal self is instantiated in the brain and neuromuscular systems - including its capability for perception, proprioception, and action - it is therefore an embodied self (see also Newen 2018). This embodied minimal self has a first-person perspective and can distinguish itself from the environment (Gallagher 2013).

Relatively, there is an elaborate debate in metaphysics regarding persons and personal identity (Olsen 2019), going back to John Locke, who defines persons as follows:

“This being premised to find wherein personal identity consists, we must consider what Person stands for, which, I think, is a thinking intelligent Being, that has reason and reflection, and can consider it self as itself, the same thinking thing in different times and places” (Locke 1979, p. 335, original italics).

There is clearly overlap between the notions of self and person, and they are sometimes used as synonyms. However, for the purpose of this article, I define person as a broader notion than minimal self. Following Locke, persons exhibit consciousness, self-consciousness, and can think, reason, and reflect about themselves and the world. These features of persons are grounded in their embodiment and, importantly, come in degrees. Because they can think,
reason, and reflect, we may refer to persons as cognitive selves. The minimal self is about pre-
reflective embodied experiences and so it lacks capacities for thinking, reasoning, and
reflecting. Having these properties is part of the definition of person but not part of the
definition of minimal self. Crucially, for Locke, persons are the same over time, when they can
remember their past actions and experiences.

Personal identity is often discussed as being about sameness over time, trying to better
understand under which conditions a person at $t_1$ and $t_2$ can be said to be the same person.
Locke provides a memory theory of sameness over time, whereas others such as Olson (1997)
provide a biological theory focussing on sameness of the body. So, personhood is about
category membership, whereas personal identity is about numerical identity and continuity of
personhood. Minimal selves, by contrast, are not continuous over time, typically only existing
for a few seconds.

The narrative self is a richer notion than the minimal self and personhood, arguing that our
self has a narrative structure. Schechtman writes that “a person’s identity is constituted by the
content of her self-narrative, and the traits, actions, and experiences included in it are, by
virtue of that inclusion, hers” (1996, p. 94). A self-narrative can be characterised as a
subjective, affective, and personal story, containing a mostly accurate chronological depiction
of a series of connected events and experiences that constitute a person’s identity and self.
This self-narrative is realised by autobiographical memories, that is, memories of the personal
past, which are organised in a narrative structure. On this view, what makes you the particular
self you are and what distinguishes you from other selves is your unique autobiographical
narrative (Schechtman 1996). So, to better understand our self, we need to focus on our
autobiographical memories and how they build up a unique narrative of our past experiences
and activities. One way to look at the difference between persons and narrative selves is as
follows. The notion of person gives an account of “what I am”, whereas the notion of a
narrative self gives an account of “who I am.” What I mean here is that personhood is a
question of category membership, whereas narrative selfhood is about characterising
someone’s identity. Some theorists argue that narratives are constitutive of selves
(Schechtman 1996), whereas others argue that we merely have a narrative sense of self
(Goldie 2012). Marya Schechtman makes an ontological claim that selves are narratives,
whereas Peter Goldie makes a phenomenological claim, namely that we experience ourselves
in terms of a narrative. Goldie refer to this as a “metaphysically light” version of the narrative
self. “The narrative sense of self is quite a simple notion. It is the sense that one has of oneself
in narrative thinking, as having a past, a present, and a future” (Goldie 2012, p. 118).

Schechtman’s constitutive narrative view doesn’t demand that we have formulated our entire
life story in detail, but when asked (by ourselves or others) we should be able to articulate
(parts of) it. The narrative should also be a mostly truthful depiction of past events and
experiences. One of her main points is that we experience and interpret present experiences
in the light of an ongoing narrative. The experience of winning the lottery is different for a
millionaire, homeless person, and someone who struggles with a gambling addiction. The
experience will be different for these people because of their different background narrative
(Schechtman 2011). Current experiences are thus interpreted and given meaning in terms of a
past narrative, which is what makes you, you. So, my narrative influences the nature of my
experiences and characterises my self and identity.
One way to think about the relation between these notions can be characterised as follows. Agents with a narrative self are necessarily a person. What I mean is that without conscious experience and the capacity for thinking and reflecting one cannot develop a narrative self. And to be a person, one must have the capacity for conscious experience. Minimal selves also have the capacity for conscious experience. Persons and minimal selves thus share this capacity. But agents with a minimal self aren't always persons and don't always have a narrative self. Some neonates, for example, presumably have a minimal self but aren't (yet) persons and don't (yet) have a narrative self (for more discussion on the relation between the minimal and narrative self, see Menary 2008; Krueger 2011). Whether an entity has a minimal self depends on whether it has the capacity for experience. There is empirical evidence to suggest that neonates have this capacity (Rochat 2011). Also, there are persons without a narrative self. Young children don't yet have a narrative self, which typically develops during adolescence, but are considered persons, as they can think, reason, and reflect to some degree.

Having this brief outline of minimal selves, persons, and narrative selves in place, we'll look at whether minimal selves, persons, and narrative selves can be extended in the next sections. Before doing so, two brief clarificatory notes are helpful. First, the word “extended” here means that the entity in question includes parts beyond the skin. So, when authors claim that the self is extended, they claim that the self (minimal, narrative, or otherwise) is not an entity realised by our biological organism, but is an entity that has both biological and artifactual parts. Those artifactual parts, then, literally become part of the self. Second, as stated above, there is a lot of debate on the self and there are many interpretations of what the word “self” refers to. As Olson (1998) argues, scholars cannot agree on the characteristic features of selves. For example, minimal selves and narrative selves are quite different types of entities. Likewise, the phrase “extended self” refers to a different type of entity in different authors. The way the authors discussed below (Clark & Chalmers, Heersmink, Wallace, Belk, Piredda & Candiotto, Howell) use the phrase “extended self” thus differs. So extended minimal selves, extended persons, and extended narrative selves are different types of entities.

3. Self and extended embodiment
Claims that technological artifacts extend human bodily and perceptual capabilities have a long tradition in philosophical theorising about technology. Marshal McLuhan, Maurice Merleau-Ponty, Martin Heidegger, Don Ihde, Andy Clark, and others have argued that some artifacts extend our motor and sensory systems. Classic examples from phenomenology include a carpenter using a hammer to act on the environment (Heidegger 1962) and a blind person using a cane to sense the environment (Merleau-Ponty 1965). Based on these examples, a distinction can be made between two kinds of body-extending tools, those that we use to act on the world and those we use to perceive the world (Brey 2000). The blind person's cane is primarily used to perceive the world, whereas a hammer primarily to act on the world. Phenomenologically, such tools can be experienced as transparent extensions of our sensory and motor system. Transparency comes in degrees and full transparency often takes a relatively long period of training. When transparency is achieved, the tool withdraws from conscious attention and is experienced as part of the body (De Preester & Tsakiris 2009).

Neuroscientists and cognitive scientists, too, argue that technology can extend motor and sensory abilities (e.g., Maravita & Iriki 2004). Particularly relevant are claims about artifacts being absorbed in our body schemas, which are subpersonal representations of the body's
size and location in space, feeding into action repertoires. Body schemas are flexible and can adjust to a changing biological body and to using new tools. So it makes good developmental and evolutionary sense for body schemas to be so flexible. When a tool is incorporated into the body schema, it transforms how we represent space. The brain represents space in three ways: (1) personal space which coincides with the biological body, (2) peripersonal space which is the space close to body, and (3) extra-personal space which is the space far from the body. Peripersonal space is important because it is where all physical interactions between the embodied agent and the environment take place (Serino 2019). When we embody tools, our personal space and sense of embodiment are extended and peripersonal space is transformed.

A classic example from neuroscience is the rubber hand illusion. During an experiment, a rubber hand is placed on a table before a participant and the biological and rubber hand are simultaneously stroked with a paint brush while the participant looks at the rubber hand. After a few minutes, the rubber hand is perceived to be replacing the real hand, in that the agent feels the hand to be part of the biological body, which is accompanied by a sense of disowning their real hand. This is caused by a correlation between visual and tactile input, creating a feeling of ownership of the rubber hand. When that happens, the rubber hand is attributed to one's embodied self, that is to say, phenomenologically, the agent perceives the hand as a real part of his or her body. This is possible because body schemas are flexible, making it possible to incorporate non-biological artifacts such as the rubber hand into the body's subpersonal representation of its size and location in space. Whilst the rubber hand is neither used to sense nor act on the environment, it does demonstrate that the sense of bodily ownership can extend to non-biological artifacts. More recent and complex examples are exoskeletons, artificial limbs, robotic hands controlled by brain-computer interfaces, and objects in virtual reality (see Schettler, Raja & Anderson 2019 for an overview). These artifacts and technologies, too, can (to varying degrees) be incorporated into their users' body schema, extending their personal space and sense of embodiment and transforming their peripersonal space. Insofar as embodiment and peripersonal space are important for sense of agency and ownership of experience, incorporated tools can be seen as extending our embodied selves. This type of extension may be referred to as phenomenological extension.

Let's look at a successful case of phenomenological extension. Biomedical engineers from the MIT Media Lab developed an agonist-antagonist myoneural interface, allowing a connection between muscles in the residuum of an amputee with a bionic leg. A first step in the procedure is to surgically connect an agonist and an antagonist muscle in the amputee's residuum such that contraction of one muscle stretches the other. The researchers point out that the agonist-antagonist muscle dynamics allows proprioceptive signals from mechanoreceptors within both muscles to be communicated to the central nervous system (Srinivasan, Diaz, Carty & Herr 2019). Electrodes are then placed on the muscles and small computers in the bionic limb use those signals to control motors in the bionic limb. This allows the amputee to not only use the limb but also to proprioceptively feel it, thereby establishing a bi-directional communication pathway between the embodied brain and an artificial limb.

In a TED talk (Herr 2019), Jim Ewing, the first patient with such a neuro-technological interface, describes his relation to the artificial limb as follows. “I didn't feel like a cyborg. I felt like I had my leg and it wasn't like I felt I was attached to the robot so much as the robot was attached
to me. The robot became part of me. It became my leg pretty quickly.” The bionic limb works so well and is so effectively integrated into his neuromuscular system that Ewing can even use it to go rock climbing. Inspired by a Dennettian (1984) view of self which gives intentional control a central place in demarcating the self, Clark argues that “by linking the conception of self to a conception of whatever matrix of factors we experience as being under our direct control, Dennett makes ample room for truly hybrid biotechnological selves (2003, p. 131). Artifacts such as bionic limbs properly integrated into our body schemas and action repertoires, over we can exert intentional control are, on this view, part of the embodied self. Whether the subject of experience is extended remains unclear in these cases, but the embodied system that feels and acts on the world is clearly extended (see also Malafouris 2008).

Self-tracking, which is the use of technology to record or track information about the human body, is also relevant in the context of this article. Current forms of self-tracking include the recording of body weight, energy levels, time usage, heart rate, body temperature, exercise patterns, GPS locations, sleep quality, diet, dreams, blood chemistry, and other variables. This information is related to metabolic, biological, and behavioural dimensions. One of the implications of self-tracking is that it provides a greater degree of awareness regarding one’s bodily states and processes (Smart, Clowes, Heersmink 2017; Smart, Heersmink, Clowes 2017). Self-tracking affords new ways in which body-related information is made available and quantified, creating information that is inaccessible by proprioception and perception. Seen as a purely epistemological exercise (i.e., as a means of getting to know the properties of the embodied self) self-tracking appears to be a smart strategy. Whilst this doesn't necessarily extend one's embodied self, it can be argued that information and knowledge about the embodied self is extended into self-tracking technologies (see also Carter & Pritchard 2019). The relation between agents and self-tracking technologies can be seen as an extended and distributed cognitive system.

4. Self and extended mind
The extended mind is the notion that minds and cognitive systems are realised by embodied agents interacting with environmental structures such as cognitive artifacts. Artifacts can become reciprocally integrated into the cognitive systems of their users and are then literally part of those systems (see Menary 2010 for an overview). The extended mind is sometimes motivated with a thought experiment: Otto, a man with Alzheimer's disease, using a notebook as a memory system. The information in the notebook plays relevantly similar functional roles in Otto's behaviour and mental economy as information usually stored in semantic memory. It's furthermore reliably there when he needs it, available to consciousness and to guide action, and the information in the notebook is trustworthy. The point here is that the notebook is not an aid to but a constitutive part of Otto's overall memory system, reciprocally integrated into Otto's memory system.

Clark and Chalmers (1998) tentatively argue that the extended mind entails an extended self.

“Does the extended mind imply an extended self? It seems so. Most of us already accept that the self outstrips the boundaries of consciousness; my dispositional beliefs, for example, constitute in some deep sense part of who I am. If so, then these boundaries may also fall beyond the skin. The information in Otto's notebook, for example, is a central part of his identity as a cognitive agent. What this comes to is that
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. Far better to take the broader view, and see agents themselves as spread into the world” (1998, original italics, p. 18).

In this quote, Clark and Chalmers reject that the self is a mere subject of experience or a bundle of occurrent states. Instead, they argue that cognitive states that we are not currently conscious of can be part of the self. More specifically, they argue that dispositional beliefs, for example those stored in a notebook, should be seen as part of the self. Essentially, Clark and Chalmers argue that dispositional beliefs in semantic memory are part of our self. These dispositional beliefs are extended, and thus the self, too, is extended. So, an extended belief system implies an extended self. One thing to note here is that Clark & Chalmers don't provide a substantive notion of self. They merely argue that extended dispositional beliefs are part of the self. Clark has later reiterated the extended self, writing that “our best tools and technologies literally become us: the human self emerges as a soft self” (2007a, p. 278, see also his 2003 and 2007b). The self that is extended here is neither a minimal self nor a narrative self, but something in between.

The supervenience base of his identity as a cognitive agent is extended. In other words, artifacts become part of the cognitive system that realises his identity as a cognitive agent. This is so because information in Otto's notebook plays relevantly similar functional roles as information usually stored in semantic memory and is deeply integrated in his memory system (Heersmink 2012, 2015). This is important because, as Wilson & Clark point out, “the right kind of coupling (one resulting in deep functional integration) is a major part of what determines the scope and bounds of an agent's cognitive apparatus” (2009, p. 68). Because Otto's memory system is extended, it means his identity as a cognitive agent is also extended.

We may refer to the self that is extended here as a cognitive self, as it is a self that can remember non-occurrent beliefs, which is a high-level cognitive process. On this view, my beliefs are part of who I am and what I believe is not just stored exclusively in the brain, but sometimes in the environment.

Following Clark & Chalmers, some theorists argue that the extended mind implies an extended self (Clark 2003, 2007a, b; Anderson 2008; Malafouris 2008; Hongladarom 2016; Piredda & Candiotto 2019; Milojevic 2020; Clowes forthcoming), whereas others deny this (Wilson 2004; Baker 2009; Olsen 2011; Wilson & Lenart 2014; Harris 2019). The earliest critique on the extended self is from Robert Wilson (2004; see also Wilson & Lenart 2014), who is a prominent defender of extended mind type approaches, but denies they entail an extended self or in his terminology an extended individual. He argues in favour of a wide realisation base of mental states and processes (Wilson 1994). On this view, the material vehicles of cognitive processes (numbers, words, symbols, etc.) can be located in the environment, outside individuals. Individuals themselves, however, are realised by organisms.

---

2 What I mean is that Clark and Chalmers don't provide or develop a theory of the self. They merely argue that self is more than a bundle of occurrent states, which is quite a thin notion.

3 A reviewer helpfully pointed out that there are also interesting links between extended cognition and the self and religious cognition. See for example Rockwell (2009), MacKenzie (2010), DeMoss (2011), Marmodoro (2011), and Krueger (2016).
“Why aren’t subjects or bearers of mental states themselves wide? The characterization of wide realizations preserves the idea that properties with such realizations are still properties of individual subjects. [...] my belief that Paris is the capital of France remains my belief even though it has a wide realization. [...] there is a basis for marking out individuals as the subjects of properties, even those properties with wide realizations. Individuals – and here, as always, our paradigms are individual people and individual organisms – are spatio-temporally bounded, relatively cohesive, unified entities that are continuous across space and time. In the actual world, it is individuals who form and maintain beliefs, experience emotions, and wonder about what will happen next” (Wilson 2004, original italics, pp. 141-142).

There is a lot packed into this quote, so let me try to unpack it. Properties of subjects such as beliefs and other cognitive states and processes have a wide realisation. So a subject’s beliefs and other cognitive states and processes can be extended beyond the biological organism. Subjects themselves, however, do not have wide realisations. Individual subjects realised by biological organisms have beliefs, experiences, and think (about the future). The subject of mental states (i.e., the entity that has the mental state) is thus an individual realised by a biological organism. This means that individuals are non-extended but are realised by a biological organism alone. Importantly, on Wilson’s view, which is contrary to Clark and Chalmers’ view, an extended belief system does not imply an extended subject or an extended self. So, a subject is not the sort of thing that can be extended, according to Wilson.

Lynne Rudder Baker (2009) rejects the extended mind thesis and even if it were true, she denies it entails an extended self. Here, I'll focus on her second point, i.e., the denial of an extended self. As Baker draws out the implications of an extended self, she worries that “we are just shifting combinations of biological and non-biological elements [...] Persons disappear into temporary hybrids. They become scattered objects, different hybrids at different times” (2009, pp. 645-647). Her point is that over time, we couple with different artifacts, creating an extended self with a different material composition, which implies there is no continuity of selfhood over time. Our self is then composed of different material elements at different times. So, for example, at t₁ our self is realised by an organism interacting with a notebook, at t₂ our self is realised by only an organism, at t₃ our self is realised by an organism interacting with a photo album, etc. The material supervenience base of selves then differs at different moments in time and so we disappear into what Baker refers to as “temporary hybrids”. Baker finds this hard to accept, because on her view there are enduring selves and these enduring selves are realised by biological bodies, not by bodies interacting with artifacts.

Baker’s second worry is concerned with the extended nature of selves. On her view, the self is defined by having a first-personal perspective, which is made possible and constituted by our biological body, including our brain. It’s important to note that, on Baker’s view, we are more than our biological body, because the persistence conditions for first-personal perspectives and biological bodies are different. The self endures as long as there are first-personal experiences, whereas the organism endures when there are certain biological functions such as metabolism. The relation between selves and bodies is constitution, not identity, and therefore they exist on different ontological levels. For Baker, a self or person, that is, an entity having a first-personal perspective is not the sort of thing that can be extended beyond the skin, because, on her view, a person is constituted by an organism that extends only as far
as the skin. She acknowledges that bodies can be extended with prostheses and other embodied tools, which may influence and transform our first-personal perspective but cannot extend it. On her view, the minimal self thus remains in the head or the body. Wilson and Baker essentially argue that the subject of experience is not extended, which is largely consistent with the analysis in section 3.

Olson (2011) attacks the extended self from an animalist view⁴. He argues that if Otto’s notebook is part of Otto as a person, then it means that persons are not biological organisms but organisms-plus-artifacts. “This has a momentous implication: Otto is not a biological organism” (Olson 2011, p. 486). This is difficult to accept for Olson who defends a view on persons, arguing that we basically are human animals realised by biological organisms. On this view, we are our biological bodies, our psychology and memories are not terribly important for our personal identity (Olson 1997).

Heersmink (2017ab; 2018) develops a different argument for an extended or distributed self, synthesising the extended mind with the narrative self. He argues that human autobiographical memory systems are essentially open to incorporate and rely on resources in our environment, in that way our memory systems are extended and distributed across embodied brains and environmental resources. Particularly relevant are evocative objects, which are human-made objects or structures that intentionally or unintentionally aid us in remembering our personal past (Turkle 2007; Heersmink 2018; Heersmink & McCarroll 2019). Examples include photos, videos, souvenirs, postcards, concert tickets, journals, books, works of art, trophies, inherited objects, and many other mementos. Interacting with such objects activate cognitive processes that retrieve the contents of personal memories into consciousness. These objects are integrated into our practices of remembering and aid us in remembering our past in a more reliable and detailed manner. In some cases, the autobiographical dependency on the object is so great that we can't remember the event without interacting with the object. This may happen when looking at photos of events occurred deep in one’s past, say, more than a few decades ago. A current trend is to record one’s activities and experiences with self-tracking or life-logging technologies. These technologies such as wearable cameras, GPS tracking devices, Fitbits, and other sensors allow one to create an elaborate database about one's past called a lifelog. Lifelogs are paradigm cases of autobiographical memory technologies that are deeply integrated into the memory systems of their users (Bell & Gemmell 2009).

Other people, too, aid us in remembering our personal past. Consider the below example from Celia Harris, Amanda Barnier, John Sutton & Paul Keil (2014), where a long-married couple tries to recall the name of the show they saw on their honeymoon more than 40 years ago. Neither of them knows the name when asked separately, but by giving each other cues,

---

⁴ Animalism is a theory of personal identity according to which human persons are animals. On this view, human persons are biological organisms: members of the primate species *Homo sapiens*. In the *Stanford Encyclopedia of Philosophy* entry on animalism, Stephan Blatti (2019) points out that “The conditions whose satisfaction is necessary and sufficient for a past or future being to be identical with a human person that exists now just are those whose satisfaction is necessary and sufficient for a past or future being to be identical with an animal that exists now.” So, on an animalist view, the persistence conditions of human persons have to do with our biology, not our psychology.
they jointly construct the answer, in that way generating an emergent memory system that knows more than the individual members.

Wife: And we went to two shows, can you remember what they were called?
Husband: We did. One was a musical, or were they both? I don’t ... no ... one.
Wife: John Hanson was in it.
Husband: Desert Song.
Wife: Desert Song, that’s it, I couldn’t remember what it was called, but yes, I knew John Hanson was in it.
Husband: Yes.

Such socially distributed memory systems are referred to as transactive memory systems (Wegner 1986). In such transactive memory systems, there is a cognitive interdependence between group members, in which case both agents rely on each other’s memory systems.

The material and social distribution of autobiographical memory systems has important consequences for our narrative self because autobiographical memories are the building blocks of our narrative. The autobiographical memory systems realising our narrative identity are distributed across our embodied brains and environmental objects and structures. The upshot of this view is that who we are as narrative selves depends on and is partly constituted by a distributed network of environmental structures. This argument can be characterised as follows.

(P1). Human selves have a narrative structure realised by autobiographical memories.
(P2). Some of our autobiographical memories are distributed across agent and environmental structures.
(P3). If human selves have a narrative structure realised by autobiographical memories and if some of our autobiographical memories are distributed, then human selves are also distributed.
(C). Therefore, human narrative selves are sometimes distributed structures.

Both Clark & Chalmers’ (1998) notion of an extended cognitive self and Heersmink’s (2017ab, 2018) notion of a distributed narrative self do not suggest that the minimal self or subject of experience is extended\(^5\). They argue that the cognitive machinery realising the self is extended, but conscious experience, on these views, remains in the head. Thus, the criticisms outlined above by Wilson and Baker, arguing that the subject of experience is internally realised, actually don’t apply to Clark & Chalmers and Heersmink, who argue that belief systems and autobiographical memory systems are extended, but not that the subject of experience is extended. However, Baker’s (2009) critique that, on these views, the self is a constantly changing mix of biological and technological components does hold water, undermining the continuity of self over time. There are two possible responses to this criticism. One, accept there is no continuity of self over time, in that selves do not persist, but are rather momentary entities. Whilst this may be true for the minimal self, there is at least some sort of continuity of persons provided by their autobiographical narrative (Schechtman

---

\(^5\) Neither do the views of Malafouris (2008), Hongladarom (2016) and Milojevic (2020).
Two, artifacts and other people can actually provide stability and continuity of the narrative self over time.

I’ll elaborate on the second point here. Our biological bodies and memory capacities change during our lifetime, sometimes rather significantly. But some of the evocative objects we keep in our homes or stored in our computers are quite durable and typically have fixed informational properties, providing long-lasting information we can lock onto at various stages in our life (Clowes 2012). Over time, we do indeed add new evocative objects to our ecology of memory objects and remove others, but many of such objects remain in our lifeworld for substantial periods of time (Silver 1996). This is, of course, particularly true for long-term transactive memory partners like one’s parents or spouse who often remain in our lives for many decades, supporting our autobiographical memories and narrative self for long periods of time. So rather than being a constantly changing mix of components, extended selves actually have components that are quite stable.

A relevantly similar view is developed by Kathleen Wallace (2019a, 2019b). Criticising psychological and animalist approaches to self and personhood as being too internalist, she develops a view where selves are spread out across the environment in which they are situated. Her view doesn’t rely on the extended mind but on feminist theories of relational autonomy but comes to a similar conclusion, namely that

“the self is relational throughout, psychologically, physically, biologically, culturally, semantically, as well as socially. Hence, the self is a network of relations” (2019a, p. 189, original italics). This “theory opens up productive ways of understanding the self as a socially constituted, embodied being that interacts with and extends into the physical and social world in which it moves, judges, perceives, experiences, cognizes, makes and acts” (2019a, p. 197).

She refers to this view as the “cumulative network model”. Cumulative, because the self is an unfolding network of relations between biological, psychological, and social traits and environmental structures. Consider an example:

“Lindsey is mother, novelist, English speaker, Irish-Catholic, feminist, professor of philosophy, automobile driver, psycho-biological organism, introverted, prone to a cheerful disposition, fearful of heights, brown-eyed, myopic, left-handed and so on. Traits are related to one another to form the network of traits that is the self, Lindsey” (2019a, p. 197).

My current self is the sum total of these relations, which determine how the network will unfold into the future. The self is thus not just a network but also a process. The traits constituting the network can change over time, one may lose a partner, learn to play an instrument, move to a new house, or develop an illness. “However, some cluster of traits constituting the integrity of the self must persist, although there may be no single trait or cluster of traits that is necessary” (2019a, p. 200). This view is, ontologically, similar to Heersmink’s distributed narrative self view, but her view is more radical than Heersmink’s, in that the network constituting the self is much larger than on Heersmink’s model, which only looks at environmental resources scaffolding autobiographical memory.
Surprisingly, there has been a parallel and separate debate on the extended self going on in marketing and consumer research which also doesn't hinge on the extended mind. In a wonderfully rich paper and inspired by a Jamesian view on self, Russell Belk (1988) argues that “objects in our possession literally can extend the self” (p. 145). To better understand our consumer behaviour, Belk argues, we need to understand that the objects we buy and own have a very intimate relation to their owner, i.e., they extend the self of their owner and for this reason we feel attached to the products we buy. Belk's approach and motivation are quite different from extended mind based approaches to the extended self, he primarily draws on literature in anthropology, but the conclusions are similar. He is interested in the type of objects that are incorporated into the self (e.g., gifts, collections, pets, other people) and what that implies for consumer behaviour (e.g. the type of gifts one buys). Note that both Wallace and Belk's notion of an extended self is quite different from the minimal or narrative self. They are not talking about an extended subject of experience or an extended narrative self. Belk's (1988, 2013) notion of an extended self has generated an enormous amount of debate and literature in marketing and consumer research, which has remained largely separate from more philosophically-inclined debates on the extended self. Cross-fertilisation between these debates would benefit both.

5. Self and extended emotion
The extended mind has recently extended into debates about emotion, consciousness, and character traits. In the next three sections, we'll look at extended emotion, extended consciousness, and extended character, and at what these notions imply for the boundaries of selves.

A number of theorists have recently conceptualised emotion as extended (Slaby 2014; Colombetti & Roberts 2015; Carter, Gordon & Palermos 2016; Krueger & Szanto 2016). A central feature of genuinely extended emotions is that an external resource needs to be reciprocally integrated into emotional states and processes. A mere causal influence on emotional states and processes (ala James 1890) will only show that emotion is embedded and situated. Consider a number of brief examples. When a musician practices with his or her guitar, there is an ongoing process of emotional self-stimulation. Playing some chords or riffs will generate certain emotional states and regulate in real-time the experiences of the musician. These induced states and experiences, in turn, will influence what the musician will play next, which will generate new emotional states, and so on. There are ongoing feedback loops between the embodied interactions with the instrument, the music, and the emotional states and processes of the musician. Similar processes may occur when sculpting, painting, drawing, or designing.

Emotions can also be socially extended. Dyads of infants-caretakers, romantic couples, or dance partners may reciprocally interact such that their (embodied) emotions become entangled. Drawing on the work of Michele Merritt (2013), Joel Krueger & Thomas Szanto (2016) write that “interpersonal off-loading of motor control in the rhythmic adaption to the movements of others in dancing not only can functionally enhance self-regulation but also generate emotional feedback loops on the fine-grained quality and depth of the very emotional experience of the involved subjects” (p. 868). Dancing can thus extend the emotional mind.
The quote from James (1890) with which this article began suggested that self is extended because objects and people can generate emotions in agents. Building on James’ insight, Giulia Piredda & Laura Candiotto (2019) develop an account of what they refer to as the affectively extended self. On this view, affectivity is seen as the means through which the self is extended. They criticise Clark & Chalmers’ (1998) notion of an extended self and Heersmink’s (2018) notion of a distributed narrative self as being too functionalist and not paying enough attention to phenomenological and affective dimensions of selfhood. “The point for us is not only that of analysing the degree to which scaffoldings regulate affectivity, but of recognizing affectivity as a fundamental vehicle for the extension of the self” (Piredda & Candiotto 2019, p. 132). Focusing on the nature of experience, they argue that embodied interactions with our affective niche extends the self, because emotions are constitutive of the self. “The affective scaffoldings and practices are in fact put into action in the experiential dimension of the self in its interaction with the world” (2019, p. 137). Self extension contributes to an ongoing process of self-construction, moving beyond a dualistic I-object relation to a more dynamical interaction with affectivity at its centre.

Emotions and affective states can be seen as both part of the minimal/embodied self but also as part of the narrative self. The content of experiences can be raw, unprocessed, pre-reflective emotions which are part of the minimal/embodied self. But emotions can also be part of the contents of autobiographical memories (Holland & Kessinger 2010). When retrieving a memory of a funeral by looking at a photo, for example, the retrieved contents are both informational and affective. So the notion of extended emotion can be used to argue for an extended minimal self and an extended narrative self. Piredda & Candiotto focus on the narrative self, arguing that our affective interactions with the world structure and feed into our narrative construct. In other words, memories of extended emotions can thus be incorporated as the building blocks of the narrative.

6. Self and extended consciousness

Some philosophers have argued that consciousness is extended beyond the embodied organism (Noë 2009; Kiverstein & Farina 2012; Ward 2012; Kirchhoff & Kiverstein 2019). The extended consciousness thesis claims that the physical machinery realising the what-it’s-like-ness of experience is extended, as to include technological artifacts. On this view, it is not just the embodied brain that generates consciousness, but an embodied brain coupled with and interacting with certain technological artifacts that generates consciousness. These claims are typically based on enactivist theories of perceptual experience (Noë 2009) and more recently also on predictive processing type approaches (Kirchhoff & Kiverstein 2019). For example, when describing the relation between agents and sensory substitution devices⁶, Alva Noë writes that “the world itself can be described as belonging to the very machinery of our own consciousness” (2009, p. 65). The extended consciousness thesis is controversial, even amongst extended mind theorists (Clark 2009; Wheeler 2015; Chalmers 2019). These philosophers argue that current arguments for the extended consciousness thesis fail and

---

⁶ These are devices that substitute the characteristics of one sensory modality into stimuli of another sensory modality. Typical sensory substitution devices contain a sensor, a coupling system, and a stimulator. For example, the images created by a camera are translated by a coupling system into soundscapes or tactile feedback, allowing a user to sense the environment without using one's eyes. Paul Bach-y-Rita, who was a pioneer in this field, showed that subjects are (to a certain degree) able to recreate an impression of their environment and experience a sense of objects in space.
that the physical machinery underpinning the what-it's-like-ness of experience is not extended. I don’t have the space to outline this debate in detail and evaluate the argument in favour and against, because to the best of my knowledge, no one in the literature has used an extended consciousness claim to argue for an extended self (yet), although Susan Hurley (1998) seems to hint at the possibility. It is, however, easy to see how this can be done. If self is a subject of experience constituted by conscious states and processes and if these states and processes are extended, then the self is extended, too. This type of argument would focus primarily on the minimal self (see Kyselo 2014 and Garavito 2019 for related discussion on enactivism and self).

7. Self and extended character

Drawing on the extended mind, philosophers have recently argued that human character traits, or more specifically our virtues and vices, can be extended (Alfano 2014; Howell 2016; Alfano & Skorburg 2017; Skorburg 2019; for related discussion see Sneddon 2011). These authors argue that the situationist critique on virtue theory suggests that our moral behaviours are so dependent on our social context that it is best to see our virtues as socially extended. A virtue is a positive character trait (or excellence) acquired through learning and habituation, that is, it is a disposition to behave in positive ways. In the *Nicomachean Ethics*, Aristotle (2004) distinguishes between two kinds of virtue, those of the intellect and those of moral character. Examples of intellectual virtues include curiosity, intellectual autonomy, open-mindedness, and attentiveness. Examples of moral virtues include courage, honesty, generosity, and justice. Linda Zagzebski points out that “A virtue is a deep quality of a person, closely identified with her selfhood” (1996, p. 104).

Empirical research in moral psychology shows our moral virtues are not stable and consistent across situations (for an overview see Doris 2002). Contextual or situational factors including ambient sounds, ambient smells, ambient light levels, mood elevators, mood depressors, social expectation signalling, and figures of authority strongly influence the way we behave. Some argue that if people in general do not show constancy of behavior predicted by attributions of virtues, then perhaps there are no virtues. Others, however, argue that we should take a systems perspective. Whether we exhibit virtuous behaviour isn’t only a function of the isolated embodied agent but of the interaction with the environment in which that agent is situated. Morality, it seems, is situated in a similar way as cognition is situated. Our inclinations, dispositions, and behaviours are so strongly influenced by situational factors that some theorists have suggested that those factors are constitutive of our virtues. In other words, virtues are extended or distributed across embodied agents and their environment, which happens when the behavioral and cognitive systems of two agents are reciprocally integrated.

Mark Alfano and Joshua Skorburg (2017) provide an example of two close friends: Azim and Ashley. They care about each other and are genuinely concerned about each other’s moral character. When one perceives a flaw in the other’s character, he or she will point that out, and vice versa. There is furthermore a strong mutual trust and a genuine concern about the other’s moral opinion. True friendships are often characterised by emotional feedback loops and can be modelled as a coupled system with strong reliable signalling and attuned receiving in both directions.
“Azim's expectations for himself, his self-knowledge, his understanding of which actions are available to him, his motivation, the reasons that appear salient to him and their weights, and his deliberative strategies – all of these are influenced in a systematic and ongoing way by Ashley. [...] Given the tight coupling and reliable feedback present in cases of friendship, we contend that friendship can be understood as a case of extended moral character” (Alfano & Skorburg 2017, p. 475).

Robert Howell (2016; for discussion see Skorburg 2019) argues that extended virtues imply an extended person, developing the following argument. Here is his reasoning:

“Virtues, vices, and other character traits can be extended, in the sense that the grounds for these traits might not be within that individual’s skin. If this is the case, persons are likewise extended in that a person can have constituents that are not within the skin [...] What we include as part of the person is determined by what persons must be if they are to bear the normative weight we give them and in particular if people are to be praised or blamed for the virtues and vices we ascribe to them” (Howell 2016, p. 147-148).

So, if character traits are a constitutive component of persons and if character traits are extended, then persons are extended, too. Note that Howell doesn’t claim that selves are extended, he argues that persons are extended, so his focus is not on the nature of experience or narratives, but on personality traits. Thus, one thing to note here is that Howell doesn’t develop a substantive notion of personhood, essentially only claiming that character traits are a constitutive component of persons, which I agree with, but it doesn't say much about the ontology of persons or selves. So, basically, he doesn't elaborate on what it is that is extended. In addition to this argument, Howell develops a set of conditions that need to be satisfied if character traits and persons are extended, focussing on the functional role of character traits.

“One role traits play is explanatory - we can appeal to them to explain why people do what they do. Another is predictive - we can predict with some reliability what people will do based on their character traits. A third is evaluative - we use character traits to normatively evaluate a person” (2016, p. 159).

The social environment thus needs to play explanatory, predictive, and evaluative roles in our moral dispositions in the right sort of way. If the social context sufficiently plays these roles, then it is part of the disposition. If it doesn't, then it isn't plausibly part of a trait and person. So, for example, when we can explain why and predict that Azim will be more generous when he is with Ashley and are able to evaluate his actions, then his moral virtue is extended by Ashley's moral character.

In a series of experiments, Nina Strohminger and Shaun Nichols (2014) have shown that many people believe that moral traits such as honesty, loyalty, and trustworthiness are the most important properties in defining a human self. More specifically, in a number of scenarios involving, for example, brain transplants, identity change brought on by pharmaceuticals, and reincarnation, Strohminger and Nichols asked participants what they thought were the most essential properties of selves. There were several categories of properties that participants could choose from, including perceptual capabilities, desires, memories, basic cognition,
personality traits, and moral traits. Participants consistently chose moral traits such as honesty, trustworthiness, loyalty, and fairness, as the most essential properties of selves. Their results indicate that folk notions of self and identity prioritise moral traits and virtues above our embodiment, autobiographical memories, and emotions. Insofar as such folk notions are on the right track, they indicate that focusing on extended moral virtues to argue for an extended self is a fruitful strategy.

8. Concluding reflection

This survey article has shown that the boundaries of selves are fluid, shifting across biological, artifactual, and sociocultural structures. If selves are constituted by our embodiment, cognition, emotion, and character traits, and if these are extended, then selves, too, are extended or distributed. None of the views discussed in this article argue that selves, either minimal, cognitive, or narrative, are always extended or distributed. There are cases when embodiment, mind, emotion, and character traits are not extended, in which case self isn't extended either. When thinking through some problem in your head, for example, or when meditating, daydreaming, or dreaming, cognitive and affective states are realised by the embodied brain alone. In these cases, the self is an internally realised embodied self. Unlike embodiment and consciousness, extendedness is thus not a necessary property of selves, but a contingent property, depending on our particular sociocultural situation. However, in order to develop a narrative self at all, we need to have extended our autobiographical memories. Tilmann Habermas, Alexa Negele & Fernanda Mayer (2010) have shown that mothers scaffold both the content of the narrative and development of narrative skills of their children. During childhood and adolescence, we learn to create and develop our narrative self by interacting with parents, caregivers, and other family members and friends. It seems difficult to develop a narrative self without having interacted with other people who scaffold our autobiographical memories. Similarly, it is likely that evocative objects such as photo albums, videos, etc. play important roles in developing a narrative self.

The implications of the extended self are phenomenological, ontological, methodological, and normative. Embodied tools and other people (e.g., dance partners) can be experienced as part of one’s self. For phenomenological and theoretical reasons, self should be conceptualised as a relational, extended, and distributed entity. To examine this entity, we should enlarge the unit of analysis in the conceptual and empirical study of the self. Philosophy should further conceptually analyse (the extended nature of and relations between) embodiment, cognition, emotions, and moral character traits, and environment and what that implies for the boundaries of selves. Neuroscience, psychology, psychiatry, and cognitive science should focus on empirically examining our embodied interactions with environmental resources and structures, instead of trying to locate the self in the brain (Feinberg & Keenan 2005).

These empirical sciences should also examine and improve disorders of the extended self (Hoffman 2016; Krueger 2020; Malafouris 2019). Krueger (2018), for example, has conceptualised schizophrenia as a disorder of the scaffolded self. Schizophrenia is characterised by an “unworlding”, which is a feeling of disconnect from lived spaces, objects, and structures. Within this unworlding experience, schizophrenia patients experience the world as alien, devoid of meaning, and objects are no longer perceived as “ready-to-hand”. Normal people construct affective niches and trust their affective scaffolds to regulate their feelings and moods, but schizophrenia patients no longer trust their affective scaffolds. Due to this lack of trust, schizophrenia patients no longer use their affective scaffolds which has
negative effects on their sense of self as they can no longer regulate their emotions. Schizophrenia patients are thus shielded off from their affective scaffolds.

On Schechtman's narrative self view, “the limits of a person are determined by the limits of a narrative, and the integrity of a single person consists in the unity of a narrative” (2012, p. 336). Thus, if memory disorders disintegrate a person's autobiographical narrative, then it also disintegrates one's narrative self. Empirical sciences can study how objects can aid patients with memory disorders in preserving their narrative identity (Miles, Fischer-Mogensen, Nielsen, Hermansen & Berntsen 2013). Given the large amount of people with memory disorders such as, for example, dementia, this seems like an important thing to do. Commenting on the effects of multimedia biographies (which are lifelogs containing photos, videos, music, and narration) for Alzheimer's patients, Masashi Crete-Nishihata et al (2012) write that “family members and participants perceived the multimedia biographies as a means for preserving the personhood of their loved one…” (2012, p. 101, italics added). There is a lot of potential for memory technologies to support the declining narrative self of patients with memory disorders like dementia, amnesia, or brain injury (Malafouris 2019), which is not yet well-understood.

The loss of evocative objects and transactive memory partners may cause a reduction in our distributed narrative self, in that we may not be able to remember parts of our narrative anymore. James wrote:

“We feel and act about certain things that are ours very much as we feel and act about ourselves. Our fame, our children, the work of our hands, may be as dear to us as our bodies are, and arouse the same feelings and the same acts of reprisal if attacked” (1890, p. 291). “Although it is true that a part of our depression at the loss of possessions is due to our feeling that we must now go without certain goods that we expected the possessions to bring in their train, yet in every case there remains, over and above this, a sense of the shrinkage of our personality, a partial conversion of ourselves to nothingness, which is a psychological phenomenon by itself” (1890, p. 293).

In line with James' view, a recent study of smartphone use shows that smartphone-users feel a reduction of their identity and some level of anxiety when separated from their smartphone (Clayton, Leshner & Almond 2015). Smartphones are paradigm examples of evocative objects and lifelogging technologies, storing a lot of autobiographical information such as photos, videos, emails, text messages, etc. and allowing communication with other people, possibly transactive memory partners. If environmental resources are part of the self, they obtain the same moral status as neurological and biological parts of the self (see also Levy 2007). This moral status of objects implies we should not interfere with and respect people's extended self. This seems particularly true for people with memory disorders who heavily rely on objects to remember their narrative, but it is also true for cognitive healthy people.

A view of selves as essentially open to incorporate objects into their actions, practices, emotions, moral character, and narratives does not imply an anything goes mentality. On the contrary, “the realization that we are soft selves, wide open to new forms of hybrid cognitive and physical being, should serve to remind us to choose our bio-technological unions very carefully, for in so doing we are choosing who and what we are” (Clark, 2007a, p. 279). On the
basis of this insight, we need to ask ourselves as individuals and as a society what kind of extended selves we want to be. Future conceptual and empirical research should address this question.

References


