Límite | ***Revista Interdisciplinaria de Filosofía y Psicología***
Límite | ***Interdisciplinary Journal of Philosophy & Psychology***
(2020) 15: 22

THE INTERNET EXTENDED PERSON: EXOSELF OR DOPPELGANGER?

Robert Clowes*
Universidade Nova de Lisboa
Lisbon-Portugal

Recibido septiembre de 2018/Received September, 2018
Aceptado septiembre de 2019/Accepted September, 2019

ABSTRACT

As the Internet becomes the pervasive background to so many of our cognitive activities, it moves beyond simply being a tool and becomes a new sort of cognitive ecology. Our deepening reliance upon it, reshapes many of our cognitive activities and this provokes profound changes in our sense of self and agency, even in who and what at we are as persons In the process we may be becoming *Internet Extended Persons*. This article uses some of the theoretical resources of 4E cognitive science to explore a central dilemma: What is the philosophical significance of these changes for us as persons? Should we view at least some Internet systems and applications as potential extensions of ourselves, both as persons and agents: as genuine extended selves, or, *Exoselves*? Or is it better to see the profiles and personalized systems as, merely appearing to contribute to our cognitive profile, but really undermining our sense of ourselves, our coherence, our agency, and perhaps ultimately our identity as persons? Might our interactions with the Internet really be creating dopplegangers rather than exoselves? This paper discusses the possibilities and constraints of the existence of exoselves and whether the Internet (or the Cloud) serves as a good substrate for extending persons.
Key Words: Exoself, Internet, Strong Agency, Reflective Transparency, 4E cognition.

SECTION 1: INTERNET EXTENDED PERSONS AND THEIR COUNTERFEITS

The framework of Situated Cognition is a central strand of 4E cognitive science that emphasizes how cognition is not sequestered away in the head but routinely exploits structures in the natural, social and artefactual worlds as an intimate part of its functioning (Robbins & Aydede, 2009). Placing an emphasis on situated cognition in relationship to the human mind moreover foregrounds how many of our distinctive capacities are intimately bound up with our abilities to create, exploit and incorporate artefacts and ambient resources into our cognitive operations (Clark, 1997; Suchman, 1987; Vygotsky, 1978). These capacities emerge, not as direct outcomes of an inner processing structures, but from deep interactions with human cultural resources, including public representational systems (Gregory, 1981), social interpretative practices (Zawidzki, 2013) but also the deep background of material culture: tools and artefacts (Malafouris, 2013). Although theoretical cognitive science has traditionally tended to downplay the role of artefacts and attendant practices in the constitution of our minds, this has recently started to change (Hutchins, 2010; Malafouris, 2008; Norman, 2000). Philosophical accounts have also started to grapple with how the human mind must be understood against the background or tools and artefacts upon which it constantly leans, or incorporates, in order to call forth its cognitive prowess (Clark, 2003; Donald, 2001; Menary, 2014; Sutton, 2010). This article focuses on attempting to understand the nature of human person against this situated background. In particular, it seeks to understand how and whether being the persons we are is now partly constituted by our reliance on a rich milieu of Internet-mediated cognitive technologies, and if this is true, what it might mean for the human condition in the early 21st century.

* Autor correspondiente / Corresponding author: robert.clowes@gmail.com

From the situated perspective, it is only through our ongoing reliance on artefacts and tools that our distinct cognitive capacities are disclosed. This perspective provides context for recent claims that the Internet is transforming not just our media but what we are, our sense of self and even our sense of reality (Floridi, 2014). As the densely integrated artefactual environment is transformed, it should be no surprise that we are transformed with them (Clowes, 2015b, 2019). And, as an ever-increasing range of our cognitive operations have become involved with artefacts that present and mediate ICT technology, and especially the Internet, our cognitive abilities and perhaps our minds may be undergoing a rapid evolution (Smart, Clowes, & Heersmink, 2017)

The focus of this article is not so much on the nature of this new cognitive ecology, or even upon how it might be changing human cognition more generally[1], but on what implications it might have for who we are. The major claim I will focus on here is that at least some systems we interact with on or through the Internet are becoming a densely integrated *exoself* which can be considered a integral part of the persons that we are, or are becoming.

The idea of an exoself is a sort of deeply integrated computational system that acts as a proper part or component of the cognitive systems that makes us who we are as individual persons. The term *exoself* was originally coined by Greg Egan in his novel *Permutation City*[2] "where it designates the sophisticated supervisory software that supports a digital mind online: it is able to provide information, monitor mental state(s), change it, and control its virtual environment as desired." (Sandberg, Forthcoming). Sandberg suggests that our current generation of ICTs, the array of smart, mobile and wearable technology, and especially those that are connected to the Internet, may already be morphing into a sort of exoself. To be clear, if the exoself formulation is correct it is not just that ICTs are becoming deeply incorporated into our cognitive abilities, but rather, various such systems are becoming proper parts of us and especially, who we are as selves or persons.

For anyone familiar with recent debates on theoretical cognitive science the claim the Internet might count as an exoself is easily heard in terms of the hypothesis of the extended mind (Clark & Chalmers, 1998). According to this hypothesis, some of our interactions with artefacts or systems in our proximal environment can become so intimate, and our reliance upon them can become such a consistent and important characteristic of our cognitive life that it is reasonable to treat them as proper parts of us. Given the way many of us now rely upon the Internet, incorporating it into an ever-increasing range of our cognitive operations it is natural to wonder whether Internet resources and systems should now count as part of our minds. In regard to the exoselves claim specifically, the questions becomes under what circumstances, if any, should the Internet, or especially some of its parts or subsystems come to count as part of anyone's extended self, or person?

The question I will focus on here is: what are the circumstances under which the Internet, or really systems of the Internet, could count as an *exoself*? I will interpret the claim in terms of the notion of persons, and especially, John Locke's famous definition of a person in psychological terms as a "thinking intelligent being that has reflection and can consider itself, the same thinking thing in different times and places." This allows us to reformulate the claim. Under what circumstances might sub-systems of the Internet contribute towards the psychological processes that make someone a person, or cognitive agent, that can recognize itself as the very same being over time? In order to get these questions off the ground, I will assume, that there is a non-identity between systems that contribute to my cognition, or my mind, and those that might be considered to contribute toward, or constitute me, or my identity as a person.[3] So I assume that there are systems that can form part of our minds in the sense intended in Clark and Chalmers' original paper, or more weakly make a cognitive contribution to our abilities, but should not be considered part of those systems that constitute us as persons. There are, on this view, systems that we can come to rely upon, which meet the conditions originally set out by Clark and Chalmers, that can come to play central roles in our cognitive lives, but for reasons we shall go on to examine, they are better regarded not as parts of ourselves, but ambient systems that make a less personal cognitive contribution. For the purposes of this discussion then I will interpret the claim about exoselves to mean something which is not part of my organisms that instantiate or make a major contribution to who I am as an individual human being, or my sense of who I am as such.

There are several ways we might consider systems as exoselves, including those that are (1) mere online avatars with which we interact with online systems or virtual worlds; (2) ICT systems with which we interact with as part of everyday life and that could be considered to play a role in determining or constituting our nature as individual persons (or exoselves proper); or (3) a sort of ersatz or fake exoself I will here term a doppelganger; a concept upon which I shall elaborate in a moment[4]. In this paper I am much less interested in the first option which has been treated extensively elsewhere (Boellstorff, 2008; Schechtman, 2012). I will not directly treat how it is we are represented in online worlds or how this may change our sense of self.[5]

I am interested in how the Internet mediated assistants, interactive systems and personalized data shadows may not only structure our everyday lives but increasingly contribute to making us the persons we are. Assuming such systems exist, I am interested in how we should think of them. When, and under what circumstances are they better considered as real extensions of us as individuals, what I will here call exoselves –or exoselves proper– and when they are best considered as potentially detrimental impostures that I shall call *doppelgangers*?

Why use the term doppelgangers? Partly because the term is already somewhat in vogue, used to characterize tools such as Fitbits, that we may already use to regulate certain activities, but can take on a "rogue" character, eventually acting in ways that we feel oppress us, or in ways we feel curtail who we really take ourselves to be (Bode & Kristensen). The term as I use it comes from the literary tradition–perhaps best represented by Robert Louis Stephenson's 1886 novel *Strange Case of Dr. Jekyll and Mr. Hyde*–of an evil twin or double that masquerades as a given person but acts with malicious intent against them and their interests[6]. If we accept the possibility that there are indeed Internet exoselves which extend us and partly constitute us, the possibility opens that there may also be related systems that, even though we rely upon them and use them as though they are a part of us, they may have a range of effects that could undermine or endanger our very nature as persons. Such ICT systems that appear to be exoselves but operate in ways that might subtly undermine our coherence as persons or agents, I shall term *doppelgangers.*

To examine these possibilities, in **Section 2**, I discuss three ways in which ICTs, more specifically Internet-based applications, may count as part of us as extended persons or agents. Each reflects a different possibility in the contemporary literature about the nature of self and personal identity, and different ways in which putative exoselves might contribute towards being proper parts of us and making us who we are. First, I look at systems that structure or partly realize our sense of ourselves through autobiographical memory and the sense of ongoing and connected consciousness that this establishes. I call them *Self-Narrative Systems*. Second, I discuss systems that constitute our skills or abilities, especially where the practice of those skills can be considered to define who we are through our skilful activities or capabilities. I call them *Situated Capability Systems.* Third, I look at systems that constitute our capacities for self-regulation and agency. I call them *Self-Regulative Systems.* If we do have Internet-extended exoselves, so I will argue, it will likely be because one or more of these three types of systems are realized through the Internet.

Section 3 places the question of extended personhood into the context of the debate around when a given device should be counted as a genuine mind-extender which constitutes part of the cognitive system itself and when it is better to see a given artefact as merely potent scaffolding that forms part of the environment of cognition (*e.g.*, Sterelny, 2010). Heersmink and Sutton (2018) have recently recast this debate by advancing a multidimensional framework of factors that allow us to consider the depth of cognitive integration in any particular system. Through careful consideration of the accounts of extended persons in the previous section, I argue that even the sophisticated multidimensional approacch developed by Heersmink and Sutton cannot in itself provide a key to which systems should be considered as genuine person-extenders. This is because, of the three accounts of persons discussed in the previous section each require different integrative relationships to be in place for a given cognitive system to be extended. What is of issue here is not in any straightforward way the density of integration between an agent and an artefact –although this is important– but the nature of the cognitive function that integration supplies to the agent. A system that putatively extends

an agent with regard to its situated capability systems will likely be considered more or less densely integrated along a different set of factors that one which extends an agent along (*e.g.*) its self-regulative or narrative-self systems.

Section 4 discusses further circumstances in which a person (a mind or agent) can incorporate resources into itself in such a way that it remains a relatively integrated whole, and when such an integration is better understood as way of interrupting or destabilizing us as stable coherent persons. The crucial feature I argue is that they remain open to a degree of what I call *agentive self-regulation* which is made possible especially where a given technology allows a degree of *reflective transparency.* It is our capacities to review, plan and submit our mental life to examination in the processes of self-scrutiny and self-determination which are vital here. Where ICT systems –especially through reflective opacity– make certain forms of self-scrutiny and self-reflection more difficult they undermine our abilities to operate autonomously. In these circumstances potential exoselves start to operate more like doppelgangers.

Finally, **Section 5** examines some of the novel properties of a particular form of the Internet what I have previously called Cloud-Tech (Clowes, 2015b). I focus here on four properties of these Cloud-Tech resources which have implications for the question of exoselves namely, *personalization, autonomy, social entanglement* and *reflective opacity*. I concentrate on the problems of incorporating resources which are personalized to us –although not necessarily customized by us– through extensive data profiling. The other relevant properties are: social entanglement through which the cognitive properties of many social media systems are structured through the various links made through social media systems; autonomy, whereby many systems operate under the control of Ais; and reflective opacity in the sense that many Internet applications are organised by mechanisms over which have little vigilance and perhaps less cognitive control. Taken together, these four properties challenge the conditions under which agentive responsibility can be taken for these putatively mind-extending cognitive resources, and thus tend to produce systems which are better characterized as doppelgangers than as genuine exoselves.

SECTION 2: THREE WAYS OF THINKING THROUGH EXTENDED PERSONS

There is something subterranean about the way we have come to incorporate Internet resources into our cognitive lives without –in most cases– making a conscious decision to do so. We have become skilled and practiced users of a cognitively potent and increasingly mobile, ubiquitous technology which we incorporate into an ever-increasing range of our cognitive operations. Our reliance upon it can easily become so deep that it progressively it becomes invisible (Norman, 1999). It often takes a network outage, or a visit to some locale where we do not have wireless access, to reveal our cognitive reliance on the Internet.

Anders Sandberg tells the story of attending a conference in Beijing in 2006, when he became aware of his habit of constantly checking Wikipedia as he wrote a paper. It was precisely because the censorship regime present in China interrupted his normally habitual and unreflective use of Wikipedia that he became aware of his reliance upon it. Might we consider that the compromise to Sandberg's ability to access a favoured tool goes beyond a personal inconvenience to –in this case transiently– compromising who he is as a person? If we take the notion of exoselves seriously, the suggestion is that we should. (Sandberg, Forthcoming).

When we do notice, the depth of our reliance on these new cognitive technologies can be startling. David Chalmers writing in 2007 argued that his iPhone has "already taken over central functions of my brain": part of his memory, his desires, his mathematical abilities, and even his daydreaming now often depend on his use of his iPhone (Chalmers, 2007). Reflecting on similar themes, one recent paper asked whether the theft of a person's smart phone might be better regarded as a crime against us as persons (assault) than against our property (simple theft) (Carter & Palermos, 2016). If stealing my smart phone should count as a sort of personal assault rather than robbery it seems reasonable to assume it is a proper part of me. Highly personalized Internet-based apps and services may already be considered to be playing central roles in our structure as selves, persons or agents (Heersmink, 2016b).

There are in fact several rather different questions around self and persistence of self that bear on the question of whether ICTs might count as exoselves. First there are questions over what

makes us one and the same person over time. What are the reasons to say that I am the same person who worked for an Internet company in the East End of London twenty years ago, or in my teens hoped to be a Jazz fusion guitarist? These are questions of *personal identity*. Second there are questions of how I am constituted, out of what parts and how these parts and their interactions form or constitute me as a relatively unified or coherent person. Let us call these questions of *self-constitution*. Thirdly, there are questions of who I take myself to be, and how I –possibly quite fallibly– think about, understand, and identify myself. I will refer to these sorts of questions as questions about the *sense of self*. Perhaps ideally the answers to these questions would strongly interrelate: my self-constitution would guarantee my personal identity over time, and perhaps both would form the basis of my sense of self. I point out these distinctions here because there are several ways in which an exoself might be considered as a proper part of me, and various ways it might function in questions about who I am and who I take myself to be. I will not go much further into how these questions interrelate except to say that we could think of them as different aspects of personhood.

Yet, given this, it is important to note here that not all the systems composing the cognitive basis of human cognition need be considered as part of our self, our subjective sense of self, or constitutive of our personhood. At an organismic level we are beings whose cells are dying and being replaced at a rapid rate and yet we generally take ourselves and our fellows as beings who exhibit psychological continuity over time. Some systems of our brains and bodies –and as we shall see, possibly more extended systems– constitute us as enduring and more or less coherent beings to a greater extent than others[7]. But which cognitive systems are those that are the foundations of this continuity?

In what follows I build upon contemporary work from the recent literature that suggests different ways that artefactual systems might contribute either to our personal identity, or to self-constitution, or our sense of self. I identify broadly three types of extended system that might contribute to one or more of these aspects of personhood. I will not directly venture a conclusion over which system or set of systems should be given priority –if indeed any should. Rather, I will look at three different approaches to what it might mean for the Internet (or parts of it) to operate as an exoself. Although these approaches are not exclusive, and each build in rather different assumptions, I will first treat each as a separate thesis about the way in which the Internet or its parts could serve as an exoself.

First is the *narrative approach to self*. This approach holds roughly that much of our sense of self and especially our sense of continuity of self is determined by our ability to view our lives in terms of a personal narrative. The approach has its origins in John Locke's idea that continuity of personhood depends upon, indeed is constituted by continuity or persistence of consciousness (Locke, 1979, p. 335). Neo-Lockeans have found it hard to make sense of what it would mean to have a continuation of consciousness and have tended to focus on the sorts of cognitive connectedness made possible by overlapping chains of psychological connectedness especially as made possible by memory (Parfit, 1984; Schechtman, 2005).

Since memory is the central factor and especially the forms of episodic memory involved in the constitution of our narrative sense of self, several theorists have pointed out that the Internet can play a role in constituting our sense of self and indeed personal identity (Clowes, 2012, 2013; Heersmink, 2016b, 2018). As more of us store digital photos and other mementos online, Internet based ICTS become ever-more significant in how we remember past events. It is natural to wonder whether some Internet systems are extending our sense of self. Narrative approaches have been especially championed and developed by Richard Heersmink in the context of Internet related memory systems (Heersmink, 2015b, 2018). Heersmink argues that Internet-based systems can implement the four properties of self-narratives: 1: self-narratives are dynamically arranged (rather than fixed) systems of past events; 2: events are chosen selectively rather than presented as an exhaustive presentation of past events; 3: a subjective interpretation of events (often from a first-person point of view); 4: events are depicted in the narrative in a causally related way (Heersmink, 2018)[8].

Heersmink argues for an *extended* interpretation of this narrative account that "personal identity can neither be reduced to psychological structures instantiated by the brain nor by biological structures instantiated by the organism, but should be

seen as an environmentally-distributed and relational construct." (Heersmink, 2016b, p. 1). He develops this approach starting with socially distributed memory systems and then extends the analysis to more technical systems including lifelogging systems[9]. Insofar as a large part of the world population now uses devices like smart-phones to take pictures (select parts of their personal narratives), view, shows and distribute these pictures (thus reviewing and organising their personal narrative), there is a strong initial argument that the Internet systems play an important role in both our sense of self and sense of continuity of self over time.

A second approach I call the situated-embodied capability approach to self, or just the *situated capability* approach. This approach emphasizes how what we are as persons, and our cognitive character, is not just given in our explicit sense of self or self-reflective processes, but also through our skills (Dreyfus & Dreyfus, 1980), situated affective states (Colombetti & Krueger, 2015), habits (Butler & Gallagher, 2018) and the host of background cognitive capabilities that constitute embodied subjectivity (Merleau-Ponty, 1962)[10]. Here I want to emphasize how many of these capabilities are evoked for human beings in the interactive domain of artefacts and artefact-centred human practice (Hutchins, 1995; Turkle, 2007). Such situated capabilities are not simple productions of our brains, or even brains and bodies, but are called-forth and enacted through our interactions with and dependence upon the world of artefacts (Malafouris, 2015).

One illustration of this idea in relation to ICTS would be an architect whose abilities to carry out her work have come to depend upon her skills as a CAD/CAM operator. On losing access to some software upon which many of her skills depend, she might feel that central capacities to think, imagine and *be* herself have similarly been compromised[11]. One could argue that such implies a compromise to our architect's sense of self in ways that are more profound that the loss of access to digitally encoded memories upon which her narrative sense of self might depend. On the situated capability approach, it is not our conscious sense of what or who we are that matters here, but our pre-reflective sense of ourselves as given in our skills and practices. A guiding thought here is that systems can become deeply embedded in our implicit and pre-conscious mechanisms of cognition and action that they come to be relied upon in just the way that brain and body rely upon its own biological parts[12]. This approach concurs with Anders Sandberg's claim that Wikipedia should count as part of his exoself. Sandberg is a researcher and writer and insofar as he is constantly drawing upon Google to look up references, structure his skilful practices of writing and editing papers, it makes sense to think of these resources as an important part of what makes Sandberg who he is. This approach accords with the more general case of the extended mind, which, as Clark has it, it also naturally shades into the idea of an extended self (Clark, 2006).

Clark's *Natural Born Cyborgs* develops a scenario to illustrate this artefactual dependence of self on personalized ICTs. Clark imagines a near-future Internet user who has been interacting with an sophisticated Internet based system –the MamboBot– since childhood. This Bot has been contributing to its user's "taste for the weird and exotic for three and a half decades" (Clark, 2003, p. 129) coming online when he was five. In Clark's scenario, the system user only started to notice his Bot has been disabled after several months in which time he had been feeling "unusually flat and uninspired for a while" (*ibid.*). This scenario nicely captures how Internet systems might contribute to the identity and sense of self of an agent even while operating in the background of consciousness. The situated capability approach holds that it is our largely unconscious or pre-reflective reliance on tools and our worldly embedding that implies that some external systems should count as partly constituting us as persons[13].

As this capability approach hinges, not upon any explicit sense of self, but upon how our sense of ourselves arises in the background, through what we can do, it has a somewhat Heideggerian flavour–I originally termed it the *cognitive hinterland approach to self.* On this view, self is not some inner essence, or even dimensions of consciousness, but disclosed through various forms of worldly interaction (Escudero, 2014; Heidegger, 1927). Such worldly interactions need to be understood in terms of our situatedness in social (Lysaker & Lysaker, 2008) and artefactual world (Olsen, 2010). Another way of coming at this is to note that much of what we are may never make it into conscious reflection. If what I am as a person is only constituted by that upon which I can consciously reflect upon, then personhood would have to leave out the immense

cognitive background upon which those conscious processes of reflection, selection and refinement must draw upon. We would be left with a sort of iceberg tip account of personhood. Although not equivalent, this view connects with a wider set of phenomenological intuitions which emphasize that it is the background of consciousness, or in what is sometimes called our *pre-reflective sense of self*[14]. Insofar as artefacts play an uneliminable role in our situated capabilities they can be considered as potential part-constituters of us as persons.

Thirdly, there is what I shall call the *Regulative Agency* approach, according to which, exoselves extend the mind's ability to reflect upon and review its own mental states and thus regulate itself. This third approach concentrates on how what persons are depends upon the distinctive structure of human agency and that has also been called *the structure of a person's will.* Its intellectual backdrop is an influential treatment by Harry Frankfurt which made the link between personhood and the ability to take control of one's own will (Frankfurt, 1971). Central to this structure is the capacity to evaluate and take pro and con attitudes towards one's own desires that we might call metacognitive. On this view it is claimed that we human beings–and any other creature we might designate as persons–do not just have beliefs and desires but have the capacity to take attitudes about those beliefs and desires. On this analysis, persons have a distinctive structure to their will which involves the capacity for reflection upon one's projects and goals that we might call metacognition. This capacity is considered central to human temporally extended agency –sometimes called *strong agency*[15]– which is defined in terms of our capacity for reflection, self-evaluation, self-regulation and organisation of ourselves with respect to live projects over time (Bratman, 2000). On the regulative agency approach to persons it is our abilities to reflect on our mental states, take attitudes towards them, and–at least potentially– regulate ourselves with respect to those attitudes which is central to our status as persons (Clowes, 2020; Frankfurt, 1971).

Let us note here that ICTs and Internet based software are already being extensively used as a form of self-regulation tool, from the regulation of emotional states with iPods (Bull, 2008), to the regulation of ourselves through a host of planning and self-regulation ICTs (Clowes, 2019; Duus, Cooray, & Page, 2018). I have developed analyses of examples using tools such as Fitbit for self-regulation and although Internet systems are often associated with the dissipation of agency (Carr, 2010) there is evidence that they can be used in ways that facilitate more positive form of self-regulation (Duus *et al*., 2018). Sandberg hints at something like this in his article when he writes that: 'exoself devices act as cybernetic regulators monitoring action and promoting "virtue", (Sandberg, Forthcoming)[16]. There are questions with respect to what Michael Bratman (2000) call *temporally extended agency* over whether such tools are set up to allow the forms of reflection taken to be crucial to the proper sort of self-regulation (I return to these questions below in Section 4).

We can ask here whether Sandberg's sense of the feeling detached from his exoself meets up with any of these factors in our account of the three sources of personhood. It is not obvious that the attenuation of his link to Wikipedia either greatly effects his narrative sense of self. Most of us are unlikely to be so reliant on the Internet for the immediate organisation of autobiographical memories (at least synchronically). It is thus unlikely Sandberg really felt that is sense of himself as distinct person with his own life narrative was seriously compromised by temporary lack of access to the Internet[17]. Neither do his abilities of self-reflection nor self-governance directly affected by the censorship regime[18]. It is more natural to associate Sandberg's concerns along with our second factor: situated capabilities. Given Sandberg's sense of self and even person identity is likely richly bound up with his status and profession as scholar and academic, and moreover those capacities depend on what the cognitive ecology made possible by the Internet, it is likely his sense of himself is somewhat compromised by interference with his access to the Internet.

Given this initial analysis I will now venture two observations. First, just because some of our cognitive systems depend upon or are even constituted by the Internet does not mean that loss or compromise to those systems will necessarily compromise our sense of self, or, who we are as persons. Second, whether such a compromise really has these implications will depend upon both what approach (or combination of approaches) we take to the extended self. On a narrative approach it will be systems which influence or constitute autobiographical memory which will

be of importance. On a capability approach it will be those systems that compose the array of ambient technologies on which particular skills or capabilities depend that will be central to our sense of self. On a regulative agency approach it will be those systems which influence our ability to plan, reflect upon our decisions and organise ourselves. This of course complicates our picture of what extended systems might count as exoselves. This will require further analysis, but at this juncture we might say that –in the absence of further accounts of how extended systems might come to count as proper parts of us as persons– we will assume it is only systems that we interact with in one of the three ways described above that are likely be constituents of who we are[19].

SECTION 3: THE PERSONAL INCORPORATION OF TOOLS

In a celebrated thought experiment from Clark and Chalmers (1998) paper, we are asked to imagine Otto, who suffers from Alzheimer's disease but has come to rely on his trusty notebook to organize his life and store the plans and memories which now tax his biological brain. Because of the functional similarities with how Otto stores and organizes the believes and desires in his notebook, we are to consider that not just Otto's cognition, but his mind is extended into, that is, partly realized by his notebook. The Extended Mind tradition gives us a set of criteria, that attempt to settle the boundaries of the mind in a way that doesn't make an arbitrary array of resources parts of our minds. The conditions are, first, **Availability**: " the notebook is a constant in Otto's life - in cases where the information in the notebook would be relevant, he will rarely take action without consulting it." Second, **Accessibility: "**the information in the notebook is directly available without difficulty." Third, **Trust:** "upon retrieving information from the notebook he automatically endorses it."[20]. The conditions have become known as "trust-and-glue".

In one of the first treatments of the question in relation to Internet resources, Paul Smart (2012) argued that web-resources circa 2012 were not well poised for incorporation of cognitive lives of users. However, as a main mode of accessing the web has since become the smartphone, these concerns have arguably lost much of their force (Clowes, 2015b). A cursory examination of today's highly mobile Internet gadgetry seems to indicate many smart-phone users employ their technology in ways that seem to meet the trust-and-glue conditions. Gadgetry like smart-phones form a constant accompaniment to many user's lives, and it seems that an ever-widening set of our cognitive operations involve some usage of these technologies. Our increasingly intimate and intense usage of mobile digital technologies such as smart-phones, iPads and Fitbits suggest that certain contemporary digital technologies easily meet the original conditions (Chalmers, 2007; Clowes, 2012). But perhaps too easily! In the context of the ubiquitous Internet, data centric applications and personalized gadgetry, the original conditions seem to lead to the notorious problem of cognitive bloat; to the point that threaten a *reductio ad absurdum* of the original claims.

According to an alternative theoretical approach, the *scaffolding approach* (Sterelny, 2010) artefacts are not, for the most part, best seen as part of any individual's mind but part of a common store of technology, skill supporting artefacts and representational systems that can be leveraged by individual agents and groups of agents. On this scaffolding approach, although many of our cognitive abilities depend on the presence of the right environmental resources[21], this does not mean that such artefacts are parts of individual human minds. Sterelny argues that a central reason we might prefer the scaffolding approach is that the original trust and glue conditions are significantly too liberal. They might include all sorts of things as supposed proper parts of our minds that might be better treated as cognitively potent systems which are nevertheless not parts of us. Consider the *London Tube Map*.

A forgetful commuter might use the tube-map in ways that imply all four of the trust and glue conditions, and yet for all of that, we might be disinclined to grant that the tube-map should count as part of her mind. Why? Because the map is a public ready-made resource used in the same or very similar sorts of ways by millions of travellers each year even though they-themselves make no active contribution to it. It might make more sense to think of such resources as a sort of *cognitive commons* (Dror & Harnad, 2008). Such resources can contribute to the cognitive capacities of many individual agents without, needing to count as a proper part of anyone's mind. (Of course, many such resources might, at

the level of property relations, be considered as owned by individuals, states or corporations, but this is a separate conceptual issue).

For Sterelny, an artefact can count as a part of an agent's mind only if that agent has customised the artefact to its own mind –what Sterelny calls ***individualization***– and crucially, then come to rely on those very customisations in its cognitive routines –what he calls ***entrenchment***. He urges that the vast majority of artefacts that make a contribution to our cognitive prowess are best considered as merely scaffolds. It is only the highly trusted, and crucially for him, the individualised and entrenched resources that we should potentially consider as parts of our minds and these are very few of the total space of cognitive scaffolds. Importantly, this view does not exclude artefacts and systems as meeting those conditions and it becomes a matter of investigation of artefact to agent relations in determining which might be candidates[22].

Yet, at least some Internet resources seem to readily meet Sterelny's extra conditions[23]. Today many Internet resources have become a series of database produced systems where information is dynamically customised to meet the preferences and profiles of individual users. Social media systems like facebook require a personalized profile (or avatar?) through which we interact with others. Indeed, the sorts of systems once called Web 2.0 are more or less defined by a degree of personalisation. Individualisation of a sort appears to be the norm.

An alternative approach proposes that we consider a multidimensional space of factors in order to assess such complex matter of integration between an agent and a resource. Richard Heersmink and John Sutton (Heersmink, 2015a; Heersmink & Sutton, 2018) suggest a series of factors drawn both from the literature I have already described and some wider discussion of *second wave* approaches to the extended mind (Menary, 2010; Sutton, 2010). These incorporate familiar categories such as trust, accessibility and availability –under the refined terminology of reliability and durability– individualization and what the call transformation (related to Sterelny's entrenchment). In addition, they suggest two types of transparency –informational and procedural[24]– which are designed to capture different ways in which an artefact and its interface can be transparent-in-use. A final category *information flow* is designed to capture how the dynamics of interaction between an agent and the resource can be one-way, two-way or reciprocal. Heersmink and Sutton argue that there is no set of necessary and sufficient conditions for counting something as part of our minds or not, rather we should consider systems as more or less tightly integrated or more or less densely incorporated[25].

Our focus here is on not just whether tool might extend our minds but extend us as selves or persons. In this regard Sterelny's entrenchment and individualisation or even Heersmink and Sutton extended dimensional approach may only be quite indirect guides to what might play a constitutive role in us as *persons.* This depends much more upon what we use these tools for. Do they play important roles in one of the cognitive systems I identified in the previous section? Do they play a role in our narrative construction of self, or the basis of skilled action or over our abilities self-reflection and determination? Even a highly entrenched and customized tool that is transparent-in-use, involves reciprocal information flows, and otherwise meets the trust and glue conditions might not play a role in our sense of self if it doesn't play a role in one of the three areas I have previously identified, *i.e.*, in terms of narrative, situated capabilities or agentive self-regulation. Now let us consider how Internet apps might function in the capability role.

Let us consider Talia, a Tour Guide and driver of a three-wheeled Taxi-cycle in her native city of Lisbon. She relies upon Google Maps to maintain a list of customized favourites that help her provide customized tours of the busy roads and changeable traffic conditions in complex city in which she lives and works. Thanks to her personalized customisations of city-maps, she is able to find her way to destinations of interest to her clients, finding the quickest, most convenient and scenic route based upon what the current traffic conditions are. Her personal driving habits regularly make use of this system, and specifically the way she has customized it to her needs and patterns of use. She has moreover fully entrenched those patterns into her working life, and she has significantly customized and personalized –or *transformed*, to use the terms favoured by Heersmink and Sutton– the apps she needs to do her job. The *Transparency-in-use* (in both the procedural and informational senses) of Talia's map system depends both upon the familiarity of the systems

interface and Talia's hard-won skills build over a long-term and ongoing reliance on the system (durability). Importantly for our first concept of personhood–the situated-embodied capability approach, Talia's core sense of self becomes invested in the capabilities she possesses in part through her cognitive incorporation of the app.

Talia is also a heavy user of various social media services in both her professional and personal life. She often takes a dozen or more photos a day uploading them to one of several social media systems as part of busy social life which she tags and circulates to friends. She frequently muses upon these photos and her various posts throughout the day, noting the comments they elicit from friends and acquaintances in her social circle. Her memory of many events is in several ways dependent upon this social media. In recent times some of social media systems seem to be getting more sophisticated. The Android operating system on her mobile phone for instances integrates closely with the photo app she uses. It frequently suggests that Talia "remember the day" which prompts her review series of photos and occasions that she has taken sometimes several years ago. She has noticed how viewing these "spontaneously" suggested photos sometimes shapes her ruminations about her life for the rest of day, and Talia wonders how much her memory is being shaped by the usage of these sorts of apps. She also spends time carefully curating the images that she keeps in the cloud and shows to others. The reciprocal two-way flow of information ensures her deep engagement with her online systems and –through them– with her circle of friends and extended circle of acquaintances. Talia's sense of herself and her personal narrative is deeply involved with, structured and consolidated by her daily interaction with her social media feeds. Her ongoing narrative sense of self is deeply bound up with her use of this technology.

Finally, Talia is a careful and committed user of self-tracking devices and software which she uses extensively to measure, reflect upon and organise her life. In addition to tracking her life with photos she uses a variety of software systems including calendaring systems, task systems, tracking systems and hardware such as her trusty Fitbit –a wearable activity tracker– to track, examine and regulate many of her activities. Such systems can provide extensive capabilities for users to monitor a vast variety of data about themselves and their ongoing activities. In this case the customisations and entrenchments are clear. Talia is a conscientious user of her Fitbit. Using the accompanying software, she shares information with friends from her wide social network with whom she competes to perform the required number of steps per day, enlisting their support and feedback to motivate her fitness regime. Talia relies on her collection of devices and software systems to regulate herself. Talia is an enthusiastic life-logger and many of the apps on her mobile phone provide durable, transparent and highly pervasive systems for self-regulation (Lupton, 2016; Swan, 2013). On the regulative agency approach it is Talia's ability to *individualize* the technologies that is of crucial importance alongside her trust in the reliability of the technologies built through long reliance on their affordances. Although transparency-in-use (both procedural and informational) is important here of greater importance is what I shall call *reflective transparency.* This is the ability to control and reflectively focus upon how whether she is meeting her goals and how well her various devices support her aims. (I will explain this idea in more detail in the next section).

It is important to note here that different factors of the interaction become salient depending upon which of the sources of personhood we are interested, *i.e.*, narrative, situated skills or regulative agency. There are not obviously any one or two factors that would allow us to characterise the deep involvement of technologies in Talia's sense of self or personhood. What is important is, as Heersmink and Sutton (2018) suggest, is a range of factors, including –in some cases– what I have called reflective transparency. Yet, simply summing factors to attempt to characterise interactions as more or less dense will not in itself help us determine when a device or system is operating as an exoself. What matters is the character of the interaction and the aspect of personhood in which we are interested. Systems involved in self-narrative have a different cognitive profile from systems involved in self-regulation which have a different profile again from situated skills. The multidimensional approach may offer the theorist a heuristic and a variety of ways of analysing agent / artefact interactions, but it cannot tell us in advance of an analysis which factors will be of prime importance. This will depend on which cognitive capacities we are investigating.

Thanks to our prior analysis of personhood, however we are in the position to say that Talia, is in large part the person she is through the use and reliance upon a heterogeneous set of Internet based apps. The life she leads and the person she is, is enabled and made possible through her ongoing use and dependence upon these highly entrenched and individualised technologies. Deprived of these technologies she would not be the same person.

SECTION 4: REFLECTIVE TRANSPARENCY AND THE LIMITS OF AGENTIVE SELF-REGULATION

I now want to consider an alternative scenario where even apparently deeply integrated technological systems do not appear to operate as extensions of an individual's person; but rather, or in virtue of the particular form of the integration, endanger or undermine the grounds that the "extended agent" is a person at all. Consider Cloud-Otto –a contemporary version of Otto from Clark and Chalmers's original thought experiment– who carries his smart-phone wherever he goes. His smart-phone is connected to his own favourite social media / data repository, and he uses the device to access, track and revise his set of extended beliefs and desires in much the same way as the original Otto did with his paper and pen.

Amongst the many apps he relies upon is one called the WeGo-Everyday-Destination app which helps Cloud-Otto both track and manage his everyday travels around his home city of New York. Cloud-Otto has customised access to his copy of the WeGo app by creating his own top ten list of places he likes to visit. The app automatically records the last time he visited each of those destinations and his own star ratings based upon how much he enjoyed his visit. WeGo also links to Otto's map application and calendar and includes a set of automatic notifications telling him where he should be going each day, along with audio instructions about how to get to his chosen destination. The WeGo system maintains this list for him sending him an email notification each day about his destination and sees him safely there.

Thanks to WeGo he visits all 10 favourite locations in New York in ordered succession. If he visits MOMA today it will fall to the bottom of the list and he will then not visit it for nine days (not including Sunday when he always goes to church with thanks to helpful WeGo reminders). In this way, Cloud-Otto can be sure to visit his own ten favourite hangouts in succession and lets WeGo remember for him the next time he should go.

Now let's imagine that unbeknownst to Cloud-Otto, WeGo has just "upgraded" its services. It sends him an email about the new terms-of-service, but he was having a busy afternoon at the Empire State Building and failed to read the message in detail. WeGo has added a social media aspect to his favourites list and given him a "free" subscription to the new WeGo basic. What this means is that –unless he opts out or pays the subscription– Cloud-Otto's favourites are now just the top-ten favourite locations of visitors to New York. WeGo *basic* provides Cloud-Otto's daily destination by selecting whichever tourist attraction got the highest star rating from visitors in the last week. This has so far always been the Empire State Building. Otto is now, unbeknownst to himself no longer visiting not his pre-selected list of favourite locales, but rather is being unwittingly driven by the commercial policies and the inscrutable collective mind to visit the same destination over and over again.

Cloud-Otto has fully entrenched WeGo into his own everyday habits and activities and he uses it without reflection as transparent equipment. The system meets both the trust and glue conditions and even Sterelny's (2010) more strenuous requirements of individualisation and entrenchment. If we were to measure Cloud-Otto's interactions with WeGo on Heersmink and Sutton's (2018) multidimensional scale we would likely conclude that the interaction was dense and therefore the app likely of constitutive relevance to his mind. Cloud-Otto trusts the WeGo resources, he has highly practiced and skilled usage of them –at least by his own lights– and he has customised and entrenched the app to his own activities. His usage moreover appears to be both informationally and procedurally transparent, and, thanks to fact that WeGo tracks and records Otto's visits the informational flow can be considered reciprocal[26]. Cloud-Otto's WeGo system could moreover be seen as operating as an exoself. We can easily imagine that Otto's sense of himself and his skilled abilities to find his way through his city depend on his ongoing interaction with his good devices.

And yet, there is something very odd about counting WeGo as properly belonging to or being a proper part of Otto when he has no apparent ability

to control them or even notice that they are operating outside his intended purposes. The reason I claim is that Cloud-Otto is no longer able to operate as a self-governing agent. This is in large part to do with the lack of *reflective transparency* of the WeGo system. I shall now explain this idea.

I use the term *reflective opacity* to specifically refer to that property of ICTs whereby their mechanisms of functioning and processing operations are either partly or entirely hidden from the user's view. Reflective transparency by contrast refers to the extent to which a given ICT system reveals, or allows the user to make visible, some of the informational mechanisms – such as some idea of the underlying algorithms– through which a given interface is produced. The way in which such reflectively transparency can be manifested will however typically be by the making available to the user a capacity to customize and selectively control what information will be available to them in situations of their choice. In this way reflective transparency will often be strongly related to the degree and ease of customizability and personalization of a given interface.

I have previously sometimes referred to this quality of reflective transparency as *cognitive penetrability* (Clowes, 2013, p. 116; 2015b), in an attempt to capture the idea that in some uses of (especially) information technology we can "see through or into" the interface in a way that lets us understand something of the algorithms that determine its functioning. However, because this term is used in a somewhat different way in other areas of the philosophical literature (*e.g.*, Siegel, 2012). I prefer now to refer to reflective transparency to hopefully capture the idea that a device is somewhat open to our cognitive scrutiny and affords a degree of reflection and ideally control. Reflective transparency can often run counter to transparency-in-use (especially procedural transparency) in the sense that a device–like Heidegger's hammer which his workman acts through in order to perform some hammering–becomes, at least for the duration of the skilled action, reflectively opaque. It is important to note however that there is some flexibility here. The very same tool might be at time more procedurally transparent and at others more reflectively transparent. This will depend on a series of factors including our orientation and skill toward the tool, how it is built and designed and also what we are trying to do with it at any given time.

The property of reflectively transparency –which appears not to be treated in, but significantly complicates, the Heersmink and Sutton multidimensional taxonomy– is of great importance here because, it is in cases where such reflective transparency is low or absent that practicing regulative agency becomes impeded. There will always be degrees of such reflective transparency; arguably informational interfaces cannot be wholly transparent with respect to their mechanisms[27]. But where reflective transparency is nearly or wholly absent, it becomes hard or impossible for a cognitive agent to maintain the appropriate levels of cognitive vigilance to its putative mental states.

Thus, to return to Cloud-Otto, it is because of the functional changes in WeGo app, and especially the highly impeded relationship of reflective transparency between him and his equipment, that it is hard to see the system as operating either in his interests or out of his own will. It is harder still to see these systems as a part of him that constitute him as a person. I suppose it is still possible to claim that the system is part of Otto's mind. However, this comes at the prices of making it difficult to see Cloud Otto as a coherent cognitive agent at all[28]. The resources of WeGo may be tightly integrated into Cloud-Otto but they have undermined his status as an agent in the process. At this point, the WeGo system, even though its design may not have been maliciously intended, is best regarded not as an exoself but a form of doppelganger.

Please note how this links to the *Regulative Agency* account of personhood I introduced in Section 2 of this paper. Human-like minds–the minds of persons–have the capacity to metacognitively take attitudes about at least some of their beliefs and desires (Frankfurt, 1971). Moreover, they have the ability to regulate themselves through processes of planning, self-reflection and future-orientation (Bratman, 2000), a process that similarly relies forms of metacognition. On these related accounts, it is through inspecting our thoughts and taking stances upon them that we can regulate ourselves. Being able to enact this sort of self-regulative agency is, or so I claim, a central part of what it means to be a person (Clowes, 2020). Minds that do not have these abilities because, *e.g.* some of their parts are obscured in such a way to interrupt such planning and reflective capabilities may, beyond

some slightly indeterminate point, cease to count as strong agents and in some senses as persons.

The structure of human agency appears to require that we can make at least some aspects of our mental life open for scrutiny such that they can be, reflected upon, identified with (or rejected), regulated and in principal controlled (Frankfurt, 1987). It is possible that many of these capacities (of strong agency) originate not in a pure internal realm but through the use of tools and our habits of interaction with them (Dennett, 1996; Luria & Vygotsky, 1992; Malafouris, 2008). In previous work on agency I have emphasized how it is our abilities to externalise and (in one sense) objectify aspects of inner life that make possible agentive reflection, and taking pro and contra attitudes with respect to our self-organisation (Clowes, 2019). Tying knots in string, making marks on paper and now–in the time of ICT technology–interacting with an increasing set of "smart" devices, human beings use artefacts to examine, organise and restructure their activities and themselves. The way we appropriate and use artefacts from our material environment to produce cognitive episodes constitutes much of our mental life (Malafouris, 2013). These capabilities may then later be internalized in a way that the external material culture is no longer needed (Vygotsky, 1986), although very often we will continue to rely on props and artefacts to structure our thoughts and activities. Indeed, the use of ICTs to embody these sorts of self-control functions already seems to be very widespread (Duus *et al*., 2018). Indeed, it is likely that the agentive character of human mental life and its relationship to how we are constituted as persons is reliant upon our use of tools and artefacts. Any situated-embodied approach that takes human agency seriously needs to come to terms with the ways in which that agency is bound up with our creation and use of our tools and devices.

The corollary of this is that any artefact or system that tends to undermine these capacities to self-monitor, self-regulate and correct our plans and habits as needed will tend to undermine our strong-agentive capacities. The danger is that putative Internet exoselves can be designed or, just tend to operate, in ways that can contribute to undermining our complex self-monitoring, reflection and ultimately self-organisation. And here it is clear that at least some ICTs may systematically mask the workings of their algorithms, making the modes of operations of many ICT systems hidden. Such cognitively or reflectively opaque systems can undermine the basis of strong human agency.

However, in apparent contradiction to this point, recent discussion around the possibility of artefacts counting as part of one's extended mind have precisely turned on the question of whether agentive scrutiny is required in order for a resource to count as part of an agent (Clark, 2015; Palermos, 2014). The problem is that our own biological resources they precisely seem to work as transparent equipment that we are not able to inspect or validate[29]. Shouldn't we, for reasons of parity expect extended systems to have the same sort of epistemic role? And yet as we have seen, where I am unable to practice the appropriate cognitive vigilance –apparently requiring reflective transparency– those opaque systems appear to be operating independently of me.[30]

Orestis Palermos offers an analysis of the tight integration of certain types of vigilance that appear to be necessary for an agent to be said to have knowledge. He writes "This sense of epistemically adequate-yet unreflective-cognitive responsibility can only be achieved by agents like us, whose intellectual capacities are appropriately interconnected such that cases where there is something wrong with the way we form our beliefs or with the beliefs themselves, we will be able to notice this and respond appropriately." p. 1934. (Palermos, 2014). While I think Palermos is on the right track, it is not only the belief formation process which is undermined by unreflective cognitive incorporation. Rather, it is the coherence and integrity of the cognitive agent itself which is endangered when that agent is unable to regulate itself with its own plans and intentions. It is this ability to exercise *regulative agency*, or so I claim, which makes it possible for us to take possession of exoselves. When this sort of regulation is impossible or seriously undermined and the agent is unable to regulate its putative exo-parts then the creation of doppelganger systems become a real possibility.

I have elsewhere argued that Epistemic Feelings can play an important role in helping us integrate even highly opaque systems if we are able to adjust to their operation over time (Clowes, 2017). For, although we may lack a conscious model of how a system operates, we may nevertheless develop an intuitive sense of how even quite opaque algorithmic systems may

operate if we are vigilant toward how it tends to prompt our actions. This will be a matter of degree. Insofar as the set of capacities that underlie strong agency –regulating ongoing activities, making and keeping resolutions, or simply introspecting on the available motivations for our actions– rely upon systems over which I have weak vigilance these systems will tend to undermine my capacities to exhibit strong agency. This is where there may be some historical discontinuity. External systems which make our ability to notice changes opaque are arguably the opposite to the norm, at least where self-regulation is concerned. Unfortunately, many ICT systems that operate as exoselves may tend to undermine this sort of tool-mediated self-vigilance. It is when such systems make possible agentive self-regulation that they may be considered candidates for being exoselves. Insofar as they undermine activities of self-scrutiny, the way is open for doppelgangers.

Many systems involved in human self-regulation are indeed extended systems. They are neither natural in a sense that makes them independent of material culture, not inner in the sense of being purely private and introspective, but they exist thanks to social or artefactual means that are often partly external and open to scrutiny. We know if we have previously violated a maxim in part because we remember telling a friend about it (and they might remind us!). We remember to take out the rubbish because we leave the bag by the door. We know we have failed to take the required number of steps today because we previously set the reminder in our Fitbit app and it alerts us. I suggest there is a continuity at work here but in general agentive self-regulation is typically an extended, culturally developed and artefactually enabled practice (Clowes, 2019)[31]. Self-regulative processes are typically not purely inner in any Cartesian sense. Rather our capacities for agentive self-regulation depend upon us being able to reflect on the reasons for at least some of our behaviours, something we might call *reflective transparency*. Very often such reflection depends upon our artefactual culture. It is, in terms Lambros Malafouris has developed, materially engaged.

In some –perhaps most– situations transparency-in-use is a central aspect of artefact agent couplings if they are to count as potential mind-extenders. This partly follows the parity principle intuition that, *e.g.*, a memory-trace retrieved from an artefact should be cognitively entertained in a way which involves *automatic endorsement* in a way that is equivalent to biological cognitive systems; memory retrieval on this view should not involve agent questioning its own memory system. (However, whether memory retrieval is always such an unquestioning process is controversial.[32]) It also partly follows from an intuition about transparency-in-use that goes back to Heidegger. On his account of hammer use the hammer becomes phenomenologically lost to the craftsman while involving in some hammering. The craftsman's attention is on the target of his hammering, not the hammer itself. Otto's trust in his notebook is in part evidenced by the fact that he uses it transparently and unreflectively. However, in the case of self-regulative agency we are concerned not so with the fluent practice of skills but with the reflective evaluation of mental states. Artefacts that can be considered to extend this sort of mental activity might need to support the requirement for a degree of transparency-in-use which will allow fluid action, but they will also require a degree of reflective transparency which allow the central functions of reflection and self-regulation. A delicate balance may need to be achieved here where we are able to exhibit some degree of control over the type of transparency available in the appropriate circumstance. In general, this might be seen as a sort of limit on one version of the situated-embodied approach to persons. It is not enough for a strong agent to be always simply lost in engaged skilful practice. Sometimes is it necessary to reflect and to take a stance on one's own mental life.

SECTION 5: CLOUD-TECH AND THE POSSIBILITY OF DOPPLEGANGERS

The interactive face of Internet technology has rapidly evolved from the static web-pages that we sat down at a desktop computer to consult, into a dynamic, ubiquitous and often highly personalized and adaptive technology –accessible from smart-phone app, or an increasing variety of portable and wearable devices– that is the almost constant accompaniment of human life. This *Cloud-Tech* as I have called it, is the new hyper-mobile, personalized and interactive face of the Internet, animating an ever-increasing variety of mobile gadgetry and upon which an increasing range of our cognitive processes depend (Clowes, 2015b). Cloud-Tech is not so much a singular resource but a new order

of tools forming an ever-available informational background, or environment in which we think, act and live. From the 4E or ecological perspective (Hutchins, 2010), we can view the Internet as a new type of cognitive ecology (Smart, Heersmink, *et al*., 2017). But does the deep cognitive integration made possible by Cloud-Tech applications mean they should be viewed as exoselves, part of the substrate that makes us the persons we are, or else as doppelgangers: systems that we can come to depend upon as though they were person-extenders but may actually undermine our agentive integrity?

In concluding this article, I want to probe further how the convolved properties of Cloud-Tech can serve to seriously undermine the ability of some artefactual interactions to operate as exoselves. The individual properties are *autonomy*, *social entanglement*, *personalization,* and especially the *reflective opacity* we have already discussed. I shall briefly discuss each property along with some of the ways they interact. It is the conjunction of these properties which mean that the Internet, while being an apparently strong ground for the construction of exoselves, in actuality, also has very strong tendencies toward the generation of doppelgangers.

The first property I call *autonomy* but might also be thought of as automaticity. It refers to the growing tendency for Internet ICTs to incorporate AI technologies, such as the semantic web, deep learning technologies and a range of data-driven algorithmic systems which do not just present information to their users structure, guide and arguably manipulate user behaviour (Russell, 2019). Such systems increasingly interact with each other as over-arching system in ways that can be very distant from direct human vigilance (Smart, Madaan, & Hall, 2018). Cloud-Tech systems are often highly autonomous in that the algorithmic mechanisms that underlie their operation are set-up to work independently from any direct human intervention and are often implemented through several types of artificial intelligence system. Such systems often attempt to shape, predict or nudge that user's ongoing cognitive activity.

The specific nature of this autonomy in Cloud-Tech systems is often highly related to the second property *social entanglement.* We can define Social Entanglement as the property of some artefactual systems such that the mechanisms by which they function are in part constituted by the interaction of multiple agents (Smart *et al*., 2018). The Internet has been highly socially entangled broadly since the advent of social media systems (Clowes, 2013, 2014). It is this integration of a user's actual social network into such systems that make possible the services they provide.

Consider again our hyper-connected tour-guide Talia. Talia's newsfeed, the lists of events, stories, and friends that she sees when she taps on the Facebook icon on her smart phone will all be different from what any of her friends see. This is because, Talia has a unique stored profile on those systems. Her newsfeed is populated by the unique data the system holds about her, including: her personal details, the history of her interactions with others through the system, information that has been mined and cross-referenced with other systems, categorized information on what type of stories she has previously shown interest, in and, information about which members of her social network she has previously spend time interacting with or simply shown interest in.

Such systems have an important cognitive dimension. Insofar as Talia relies upon socially entangled and autonomous social media systems such as *Google Photos* in order structure her recall of events such systems appear to be playing a constitutive role in her memory (Heersmink, 2018). Apps such as Google Photos includes a function to "remember the day". This might autonomously suggest looking at some photos taken five years ago on the same day of the year. However, the images autosuggested will tend to be those that have garnered attention previously, *i.e.*, "liked" or otherwise interacted with either by the person who originally took the image, or, more likely, the larger social network of that user. Entanglement and autonomy can thus be seen as a sort of cognitive bias, or put differently, part of the new extended infrastructure of memory and recollection[33]. In fact, the user's social network has become part of the cognitive system by which a given event is either recollected or forgotten. The socially entangled Internet thus not only changes the character of cognitive functions refracted through it but becomes part of the mechanism of a cognitive functions such as recall. Yet, this does not in itself alienate this system from Talia.

The algorithms: predictive systems and AI routines that populate Talia's Facebook newsfeed are designed to –among other purposes– *personalize* the experience for her based upon her particular

history of interactions. We might consider that Talia has, in Sterelny's terms, significantly *individualized* facebook –or a significant subset of its systems and functions– to herself. Insofar as she has heavily entrenched the usage of this personalized system by, for instance, keeping track of her friends and keeping up on local events, Talia's use of Facebook, should therefore be a candidate as part of Talia's extended mind. In virtue of fulfilling the role of extending her narrative sense of self and who she is as a person, it should also count as an exoself. Yet there are reasons to question this interpretation.

Talia's newsfeed may be indeed highly personalized to her in the sense that it is produced for her based upon a record of her unique interactive history. Yet, as Sterelny's original examples suggest, a degree of voluntary or intentional customization may be required to play the role of individualization as originally conceived. A cook sharpens and hones his knives in a way that aids his particular cooking practice. Presumably he does so with the more or less conscious intention of facilitating cooking in a relevant way. The customization was conscious. This may or may not be the case with Talia's use of social media systems and will depend on her patterns of interaction. Where the systems have autonomously and opaquely *customized themselves* to her in ways she has not noticed or cannot control her capacities for agentive self-regulation may be compromised. Moreover, Talia's newsfeed is produced by algorithms which are highly socially entangled in the sense that what she sees is determined by invisible interactions with her broader social network and how that network interacts with Talia's profile. This can produce another form of reflective opacity. The algorithms of Google Photos tag as important those that Talia's friends have taken notice of and will later suggest that she "remembers the days", *i.e.*, direct her attention toward certain snaps made on that day of the year in the past, based upon the amount of network activity that a given image might garner. Such services are often autonomously provided by the system and is not something that Talia has necessary signed up to or has even the capacity to consciously reflected on.

In his paper on *Distributed Cognition and Distributed Morality* Richard Heersmink argues that we "should not interfere with people's distributed minds and selves." (Heersmink, 2016a). However, one problem with highly autonomous and entangled technologies, is that it may become difficult to manage and deal with one's own set of distributed systems without , albeit unconsciously interfering with the minds and selves of others. In highly socially entangled systems, it might prove difficult for one user to interact with content which will subsequently be flagged for her attention through the system, without also (often unconsciously) triggering changes in how other users view and interact with their own content. Such unintentional effects appear to be the price of using complex, socially entangled algorithmic systems to structure what content we encounter through social media systems. Insofar as social networks provide some self-representing or self-structuring cognitive service, a certain degree of "interference" with the distributed minds and selves of others may be unavoidable. This raises the question of where and when social interaction becomes cognitive interference?

There is little doubt that internet applications can be used to intentionally bias the cognitive states of their users. This has recently been experimentally shown in one highly controversial recent study where alterations in a facebook algorithm has been shown to alter the mood of its users (Kramer, Guillory, & Hancock, 2014). Such experiments reveal the potential for regulating our emotional life when interventions are made with explicit intent, but it is already clear that any algorithm that controls what posts we are exposed to is having similar influences on our behaviour and emotional states. A degree of interference seems to be necessitated by these processes. The question is whether and when we should regard such interference as having negative impacts on the individual person or subject. Many Cloud-Tech systems already embody high degrees of autonomy, reflective opacity and social entanglement. These properties determine the characteristics of our interactions with many Internet systems.

As our vital and perhaps self-constituative cognitive systems come to depend upon such highly entangled social tools, we are hardly able to refrain from interfering in each other's minds and selves. But this leaves unclear what importance we should give to the independence of our own cognitive processes. The real problem here, is with how technological systems can tend to render processes invisible that we need to be able to access, in order to regulate our mental states through distinctively human agentive practices. Where the inner workings of systems are made highly opaque it can become difficult for an agent

to gain a degree of cognitive control over what are putatively her own resources.

This returns us to a property that I call *reflective opacity.* Reflective opacity, as we have seen is the obverse of reflective transparency. It refers to our inability to easily discern or access the way that some mechanism, processing technology or algorithm works in order to produce its outcome. Interfaces to ICTs which are reflectively opaque do not allow their users to easily determine the algorithmic or mechanistic basis of how they work or indeed the reasons they work as they do. Very often, reflectively opaque technology may often be transparent-in-use although this relationship is not obviously necessitated. Insofar as an ICT is procedurally transparent but reflectively opaque, they may tempt us into deep interactions that may, in some circumstances undermine our own integrity.

The personalization of many social media systems makes use of large datasets and relies on algorithmic mechanisms that function according to principles and purposes which are typically not open to scrutiny (Clowes, 2013, 2015b). There could be several reasons for this. One might be that underlying algorithms are too complex to be easily understood. A second is that they may be proprietary and protected by (*e.g.*,) patents as they are the intellectual property of some corporation. A third (non-exclusive) reason is that the interface to a given technology may have been contrived in order to hide the manipulative nature of many algorithms that generate content on many Internet based applications[34]. From the point of view of the skilful action, such opacity is not necessary, and for at least certain purposes, might not even be desirable. Indeed the functioning of many systems is so opaque to their users that there is a little sense that we can easily track whether such systems are operating in our interests or not (Clowes, 2015b).[35]

The question of whether Internet based ICTs can ever be considered to extend us as persons turns on intuitions deriving from Bratman's account of strong agency (discussed in the previous section). It is precisely when our cognitive processes are open to our capacities to form judgments on them that we can incorporate them into us. Insofar as such processes become ever more obscured or determined by systems we neither understand nor control we lose those capacities to regulate ourselves with respect to those informational sources. Although the Internet can support the properties that would allow it to operate as an exoself, there are also deep tendencies towards the creation of *doppelgangers*[36]. Systems which are highly entangled, very autonomous and transparent enough that we neither notice their activities nor have the capacity to shape them, may be better regarded not as extending an agent, but rather act as outside influences on the agent's will. Does this mean that the Internet technologies are more likely to create doppelgangers than extend us? Although the factors of social entanglement and autonomy are undoubtedly of great importance, it is our ability to regulate and take control of the technology which is the real limit on them acting as exoselves proper.

ACKNOWLEDGEMENTS

Early versions and ideas from this paper were presented at several conferences especially the 2015 conference organised in Warsaw by Avante on Situating Cognition: Agency, Affect, and Extension and later some were further tested at the 2019 workshop on "Self and Knowledge through the Internet" organised with the Applied Epistemology Research Group at Universidad Autónoma de Madrid. Support for my personal research was given by the Portuguese science foundation FCT (SFRH/BPD/70440/2010) and also from my current research funding from FCSH Nova University of Lisbon (DL 57/2016/CP1453/CT0021). I would like to thank Shaun Gallagher, Marcin Milkowski, Kenneth Aizawa, Paul Smart, Klaus Gartner, Jesús Vega Encabo and two anonymous reviewers for their insightful comments on the draft or at presentations of the paper and especially to Gloria Andrada for detailed discussion of several drafts that have helped me to greatly improve the final paper. I would also like to express my deep gratitude to the editors of this special issue Manuel Heras Escribano and Lorena Lobo for their forbearance and support.

REFERENCES

Andrada, G. (2019). Mind the notebook. *Synthese*, 1-20.

Andrada, G. (Forthcoming). Epistemic complementarity: steps to a new extended epistemology. In R. W. Clowes, K. Gartner, & I. Hípolito (Eds.), *The Mind-Technology Problem - Investigating Minds, Selves and 21st Century Artifacts*: Springer.

Bode, M., & Kristensen, D. B. The digital doppelgänger within. A study on self-tracking and the quantified self movement.

Boellstorff, T. (2008). *Coming of age in Second Life: An anthropologist explores the virtually human*: Princeton Univ Pr.

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

Bull, M. (2008). *Sound moves: iPod culture and urban experience*: Routledge New York, NY, 10001.

Butler, M. G., & Gallagher, S. (2018). Habits and the Diachronic Structure of the Self *The Realizations of the Self* (pp. 47-63): Springer.

Carr, N. (2010). *The Shallows: How the internet is changing the way we think, read and remember*. London: Atlantic Books.

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.

Chalmers, D. (2007). Forward to Supersizing the Mind *Supersizing the Mind: Embodiment, Action and Cognitive Extension*. Oxford: Oxford University Press.

Clark, A. (1997). *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: The MIT Press.

Clark, A. (2003). *Natural Born Cyborgs: Minds, Technologies and the Future of Human Intelligence.* New York: Oxford University Press.

Clark, A. (2006). Soft selves and ecological control. In D. Spurrett, D. Ross, H. Kincaid, & L. Stephens (Eds.), *Distributed Cognition and the Will*. Camb. MA: MIT Press.

Clark, A. (2010). Memento's Revenge: Objections and Replies to the Extended Mind. In R. Menary (Ed.), *Extended Mind*.

Clark, A. (2015). What 'Extended Me' knows. *Synthese*, 1-19. doi:10.1007/s11229-015-0719-z

Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis, 58*, 10-23.

Clowes, R. W. (2012). Hybrid Memory, Cognitive Technology and Self. In Y. Erdin & M. Bishop (Eds.), *Proceedings of AISB/IACAP World Congress 2012*.

Clowes, R. W. (2013). The cognitive integration of E-memory. *Review of Philosophy and Psychology*(4), 107-133.

Clowes, R. W. (2014). Faceache: WEB 2.0, Safety Culture, The End of Intimacy and the Implosion of Private Life. In F. Negro (Ed.), *Público Privado, o deslizar de uma fronteira* (pp. 279-298). Lisbon, Portugal.

Clowes, R. W. (2015a). The Reality of the Virtual Self as Interface to the Social World. In J. Fonseca & J. Gonçalves (Eds.), *Philosophical Perspectives on Self* (pp. 221-276). Lisbon: Peter Lang.

Clowes, R. W. (2015b). Thinking in the cloud: The Cognitive Incorporation of Cloud-Based Technology. *Philosophy and Technology, 28, Issue 2,*(2), 261-296.

Clowes, R. W. (2017). Extended Memory. In S. Bernecker & K. Michaelian (Eds.), *Routledge Handbook on the Philosophy of Memory* (pp. 243-255). Abingdon, Oxford: Routledge.

Clowes, R. W. (2018a). Rethinking the ipseity disturbance theory of schizophrenia through predictive processing. In I. Hipólito, J. Gonçalves, & J. G. Pereira (Eds.), *Schizophrenia and Common Sense* (pp. 113-136). Cham, Switzerland: Springer.

Clowes, R. W. (2018b). Screen Reading and the Creation of New Cognitive Ecologies. *AI & Society: Journal of Knowledge, Culture and Communication*.

Clowes, R. W. (2018 Online First). Immaterial Engagement: Human Agency within the Cognitive Ecology of the Internet. *Phenomenology and Cognitive Science*.

Clowes, R. W. (2019). Immaterial engagement: human agency and the cognitive ecology of the internet. *Phenomenology and the Cognitive Sciences, 18*(1), 259-279. doi:10.1007/s11097-018-9560-4

Clowes, R. W. (2020). Breaking the Code: Strong Agency and Becoming a Person. In T. Shanahan & P. R. Smart (Eds.), *Blade Runner 2049: A Philosophical Exploration.* (pp. 108-126). Abingdon, Oxon, UK.: Routledge.

Clowes, R. W., & Gärtner, K. (2018 Online First). The Pre-Reflective Situational Self. *Topoi*. doi:https://doi.org/10.1007/s11245-018-9598-5

Colombetti, G., & Krueger, J. (2015). Scaffoldings of the affective mind. *Philosophical Psychology, 28*(8), 1157-1176.

Dennett, D. C. (1996). *Kinds of Minds: Towards an Understanding of Consciousness*: Phoenix Books.

Donald, M. (2001). *A Mind So Rare: The Evolution of Human Consciousness*. New York / London: W. W. Norton & Company.

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Retrieved from

Dror, I. E., & Harnad, S. (2008). Offloading cognition onto cognitive technology. In I. E. Dror & S. Harnad (Eds.), *Cognition Distributed: How Cognitive Technology Extends Our Minds* (pp. 1-23). Amsterdam: John Benjamins Publishing.

Duus, R., Cooray, M., & Page, N. C. (2018). Exploring Human-Tech Hybridity at the Intersection of Extended Cognition and Distributed Agency: A Focus on Self-Tracking Devices. *Frontiers in Psychology, 9*(1432). doi:10.3389/fpsyg.2018.01432

Escudero, J. A. (2014). Heidegger on selfhood. *American International Journal of Contemporary Research, 4*(2), 6-17.

Floridi, L. (2014). *The fourth revolution: How the infosphere is reshaping human reality*: OUP Oxford.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy, 68*(1), 5-20.

Frankfurt, H. G. (1987). Identification and wholeheartedness.

Gregory, R. L. (1981). *Mind in science: A history of explanations in psychology.* Cambridge: Cambridge University Press.

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

Heersmink, R. (2015b). Extended mind and cognitive enhancement: moral aspects of cognitive artifacts. *Phenomenology and the Cognitive Sciences*, 1-16.

Heersmink, R. (2016a). Distributed cognition and distributed morality: Agency, artifacts and systems. *Science and engineering ethics*, 1-18.

Heersmink, R. (2016b). Distributed selves: personal identity and extended memory systems. *Synthese*, 1-17.

Heersmink, R. (2018). The narrative self, distributed memory, and evocative objects. *Philosophical Studies, 175*(8), 1829-1849.

Heersmink, R., & Sutton, J. (2018). Cognition and the Web: Extended, transactive, or scaffolded? *Erkenntnis*, 1-26.

Heidegger, M. (1927). *Being and Time*. Oxford: Basil, Blackwell.

Hogg, J. (2001). *The private memoirs and confessions of a justified sinner*: Broadview Press.

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

Hutchins, E. (2010). Cognitive ecology. *Topics in Cognitive Science, 2*(4), 705-715.

Kind, A. (2015). *Persons and personal identity*: John Wiley & Sons.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, 201320040.

Locke, J. (1979). *An essay concerning human understanding*. Oxford: Clarendon Press: Oxford.

Lupton, D. (2016). *The quantified self*: John Wiley & Sons.

Luria, A. R., & Vygotsky, L. S. (1992). *Ape, Primitive Man and Child: Essays in the History of Behaviour*: Simon and Schuster.

Lysaker, P. H., & Lysaker, J. T. (2008). *Schizophrenia and the fate of the self*: Oxford University Press, USA.

Malafouris, L. (2008). At the potter's wheel: An argument for material agency *Material agency* (pp. 19-36): Springer.

Malafouris, L. (2013). *How Things Shape the Mind: A Theory of Material Engagement*. Cambridge, MA, U.S.A: MIT Press.

Malafouris, L. (2015). Metaplasticity and the primacy of material engagement. *Time and Mind, 8*(4), 351-371.

Menary, R. (2010). Cognitive integration and the extended mind. In R. Menary (Ed.), *The extended mind* (pp. 227-244). London, England: Bradford Book, MIT Press.

Menary, R. (2014). Neural Plasticity, Neuronal Recycling and Niche Construction. *Mind & Language, 29*(3), 286-303.

Merleau-Ponty. (1962). *Phenomenology of Perception*: Routledge.

Metzinger, T. (2004). *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: Bradford Book.

Michaelian, K. (2012). Is external memory memory? Biological memory and extended mind. *Consciousness and Cognition, 21*(3), 1154-1165.

Neisser, U. (1988). Five kinds of self knowledge. *Philosophical Psychology, 1*, 35-39.

Norman, D. A. (1999). *The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution*: MIT press.

Norman, D. A. (2000). *Things that make us smart*: Perseue Books.

Olsen, B. (2010). *In defense of things: archaeology and the ontology of objects*: Rowman Altamira.

Palermos, S. O. (2014). Knowledge and cognitive integration. *Synthese, 191*(8), 1931-1951.

Parfit, D. (1984). *Reasons and persons*: OUP Oxford.

Robbins, P., & Aydede, M. (2009). A short primer on situated cognition. *The Cambridge handbook of situated cognition*, 3-10.

Russell, S. J. (2019). *Human compatible: Artificial intelligence and the problem of control*: Penguin Audio.

Sacks, O. (1985). The Man Who Mistook His Wife for a Hat: Picador.

Sandberg, A. (Ed.) (Forthcoming). *Post-Human Design: The Crafted Human Body and the Exoself*.

Sartre, J.-P. (1967). Consciousness of self and knowledge of self. *Readings in existential phenomenology*, 113-142.

Schechtman, M. (2005). Personal identity and the past. *Philosophy, Psychiatry, & Psychology, 12*(1), 9-22.

Schechtman, M. (2012). The story of my (second) life: Virtual worlds and narrative identity. *Philosophy & Technology, 25*(3), 329-343.

Siegel, S. (2012). Cognitive penetrability and perceptual justification. *Noûs, 46*(2), 201-222.

Smart, P. R. (2012). The Web-Extended Mind. *Metaphilosophy, 43*(4), 446-463.

Smart, P. R., Clowes, R. W., & Heersmink, R. (2017). Minds Online: The Interface between Web Science, Cognitive Science and the Philosophy of Mind. *Foundations and Trends in Web Science, 6*(1-2), 1-232. doi:http://dx.doi.org/10.1561/1800000026

Smart, P. R., Heersmink, R., & Clowes, R. W. (2017). The Cognitive Ecology of the The Internet. In S. J. Cowley & F. Vallée-Tourangeau (Eds.), *Cognition Beyond the Brain, 2nd Edition* (pp. 251-282): Springer.

Smart, P. R., Madaan, A., & Hall, W. (2018). Where the smart things are: social machines and the Internet of Things. *Phenomenology and the Cognitive Sciences*, 1-25.

Sterelny, K. (2010). Minds: extended or scaffolded? *Phenomenology and the Cognitive Sciences, 9*(4), 465-481.

Suchman, L. (1987). *Plans and Situated Action*. Cambridge: Cambridge University Press.

Sutton, J. (2010). Exograms and interdisciplinarity: history, the extended mind, and the civilizing process. In R. Menary (Ed.), *The extended mind* (pp. 189-225). London, England: Bradford Book, MIT Press.

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data, 1*(2), 85-99.

Turkle, S. (1985). *The Second Self*: Simon & Schuster.

Turkle, S. (2007). *Evocative objects: Things we think with*: The MIT Press.

Turkle, S. (2011). *Alone Together: Why We Expect More From Technology and Less from Each Other*. New York: Basic Books.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge Mass: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and Language* (A. Kozulin, Ed. and Trans 2nd ed.). Cambridge: MIT Press.

Ward, D. (2019). Moving Stories: Agency, Emotion and Practical Rationality *The Value of Emotions for Knowledge* (pp. 145-176): Springer.

Wood, D., Bruner, J. S., & Ross, G. (1976). The Role of Tutoring in Problem Solving. *Journal of Child Psychology And Psychiatry, 17*, 89-100.

Zahavi, D. (2005). *Subjectivity and Selfhood: Investigating the First-Person Perspective.* Cambridge, MA: The MIT Press.

Zahavi, D. (2017). Thin, thinner, thinnest: Defining the minimal self. *Embodiment, Enaction, and Culture: Investigating the Constitution of the Shared World*, 193-199.

Zawidzki, T. W. (2013). *Mindshaping: A new framework for understanding human social cognition*: MIT Press.

NOTES

1 These are questions I and colleagues have considered in a series of publications (Clowes, 2013, 2015b, 2017, 2018b, 2018 Online First; Smart, Clowes, *et al*., 2017; Smart, Heersmink, *et al*., 2017). I do here however return especially to the implications of ubiquitous Internet technology in the final sections of this paper.

2 Anders Sandberg traces the history of the term in his forthcoming article *Post-Human Design: The Crafted Human Body and the Exoself (Sandberg, Forthcoming)*. Sandberg also defines the term exoself on his *Transhumanist Website* in the following way: "Systems linked to the self in a cooperative way, extending the mind and the body. Especially used about the systems supporting an uploaded personality, providing information, virtual reality and monitoring." https://www.aleph.se/Trans/Words/e.html#EXOSELF

3 I will use the terms person and self somewhat interchangeably in this paper. The term self is notoriously ambiguous. Many theorists belief the term refers to many different systems coming online in human beings at different stages of developmental history (Neisser, 1988). The terms person is useful because there seems to be a little more rigger in its use.

4 It is worth emphasizing here that these three possibilities are not necessarily exclusive. For instance Turkle emphasizes in earlier work emphasizes how avatars can operate as an self-extensions (Turkle, 1985) and in more recent work how our avatar's can act to constrain and limit us in various ways (Turkle, 2011). It is possible that a system might start our as being an avatar, become an exoself and then (because of some change in the algorithms underlying the system) function as a doppelganger.

5 See Schechtman (2012) for an argument of how the avatars that some people inhabit in online worlds should, in some circumstance, be considered part of the people through which they interact. In fact, it seems likely that online avatars and profiles are rapidly morphing into systems that regulate our lives in "real life". Whether we should consider these systems proper parts of ourselves as individual agents, or as outside forces bringing us under their control is a question I will grapple with in this paper. Luciano Floridi (2014) has recently argued that the distinction between the online and virtual and offline is becoming of ever less practical relevance.

6 Although my primary example here is *Dr. Jekyll and Mr. Hyde,* which, because of its fame is likely better known to the reader, a more exact but less well known literary example is James Hogg's 1824 *Confessions of a Justified Sinner* (Hogg, 2001). Mary Shelley's 1818 Frankenstein can be seen as another (earlier) variation of the doppelganger. This theme which has deep roots in world literature.

7 Indeed, as I will argue below, some of the systems that constitute us human beings are as much a part of the social and technical niches we inhabit as parts of our brains and bodies.

8 For an illuminating discussion of self-narratives and the four properties mentioned see *Persons and Personal Identity* (Kind, 2015).

9 I discuss lifelogging systems extensively here (Clowes, 2013).

10 I have developed a –non-narrative– version of the socially situated approach to self with respect to philosophical psychiatry in several publications (Clowes, 2015a, 2018a; Clowes & Gärtner, 2018 Online First). My approach especially builds upon work analyzing some compromise to the sense of self that takes place in early schizophrenia (Lysaker & Lysaker, 2008). This view turns on the idea that at least some part of our pre-reflective sense of self is given through the ability to attune to social situations.

11 The sense of personal injury here is perhaps more closely connected to the old Marxist idea that workers cannot truly be liberated until they own the means of production.

12 For a detailed investigation of these sort of processes around the practices of a potter at the wheel see Malafouris (2008).

13 It is worth noting the differences between Clark's MamboBot and Sandberg's use of Wikipedia in his writing. The Mambot is more active and autonomous, finding and feeding information that effects its user's choices in the background of consciousness. It has also been (albeit passively) highly customized to its user and some of his cognitive capabilities have come to rely upon it. Sandberg may not notice he is constantly accessing the web as part of his writing activity but his use of the system is more active and potentially open to consciousness.

14 While using the term pre-reflective sense of self (Sartre, 1967) here I do not intend it to be understood in quite the traditional sense of a thin, or essentially contentless form of pre-reflective self-awareness which as recently and influencially been developed by Dan Zahavi (Zahavi, 2005, 2017). Rather the notion I have in mind is of a somewhat thicker but still pre-reflective sense of self that might be grounded in the notions of habit, skills or a pre-reflective but immersed sense of oneself as an inhabitant of customary social situations (Clowes & Gärtner, 2018 Online First).

15 The concept of *strong agency* –sometimes known as temporally extended agency– was originally developed by Michael Bratman (2000). I have recently developed account of the artefactual dependence of strong agency and especially its relationship to Internet mediated ICTs in (Clowes, 2019). I discuss the concept of strong agency and its relationship to personhood in (Clowes, 2020).

16 Sandberg is clearly aware that such systems may not however just promote 'virtue', for he immediately goes on to write that such regulative systems "can easily backfire when the full existential and social context is not taken into account."

17 As a sort of reference point here we might think of here of Oliver Sacks patient William Thompson or as Sacks dubs him, "The Lost Mariner" (Sacks, 1985). Mr. Thompson was a patient with Korsakoff syndrome. One striking thing about this compromise to his narrative sense of self was that until pushed the patient did not realize his deficits. Our unconscious minds are inveterate confabulators, forever filling in, missing details. There remains the possibility then that just because we do not notice compromise to our narrative sense of self, when we are lose connection to our distributed media, there might nevertheless be a compromise.

18 There is some debate here about if he were to move to China. Long-term residence under a censorious regime might indeed prevent him from looking at his photo-albums, if they were stored on google. He might also worry that some of his projects were being surveilled by the Chinese State and therefore refrain from using them.

19 A further complicating factor is whether these different factors of personhood might in fact reduce to fewer underlying factors. Unfortunately this question goes beyond the scope of this paper but the curious reader might investigate a related line of thought here (Ward, 2019).

20 In fact, in the original paper there were four criteria, Trust being described in terms of **Automatic Endorsement** as described in the main body of the text and **Past Endorsement**, namely that "the information in the notebook has been consciously endorsed at some point in the past, and indeed is there as a consequence of this endorsement." (Clark & Chalmers, 1998). The criteria as expressed here follow the Clark (2010, p. 53) presentation of these ideas and the discussion in Smart, Clowes, *et al*. (2017, pp. 52-53).

21 The term scaffolding can sometimes cause confusion because it suggests a scaffold is something which is used to construct a particular skill and can then be discarded. Indeed the literature in cognitive science was first used to indicate supports for skill development in an approach to developmental psychology (Wood, Bruner, & Ross, 1976). However, although there seems to be some influence here this is not the use of the term that Sterelny seems to suppose. For Sterelny a scaffold is a cognitive support that does not only set up a skill but has enduring potency in its maintenance and use.

22 See Heersmink (2015a) for a detailed treatment of some of the factors that might effect whether we see any given artefact to agent liaison as counting as part of the agent's mind as opposed to "merely" counting as part of that agent's cognitively potent environment.

23 See Clowes (2015b) for a more detailed discussion of these themes.

24 For further discussion of the informational and procedural transparency the reader should refer to (Heersmink, 2015a; Heersmink & Sutton, 2018). However, I do not use quite the same terminology in this paper, and instead mainly continue to use the term transparency-in-use (Clowes, 2013) in a way that subsumes Heersmink's procedural transparency but also does not exclude informational aspects. I am not convinced procedural transparency can be easily distinguished from informational transparency in many practical contexts. An ICT interface might incorporate both formal and practical aspects in its design and transparency-in-use requires being fluent with both insofar as they are separable. There may be contexts in which it is helpful to decompose these and contexts where it is not. In the next section I discuss how these factors of transparency-in-use should be contrasted with what I call *reflective transparency.*

25 Unfortunately a full analysis of this dimensional approach goes beyond the scope of this paper, but for a extended analysis using its terms, see (Smart, Clowes, *et al*., 2017, pp. 65-69)

26 Arguably there is a sticking point here. Cloud Otto is not in fact receiving some crucial information from his device namely that it is no longer updating its list of destinations in the way he expects. Nevertheless, there is quite a bit of bi-directional information flow in this example. Cloud-Otto smart phone is continually updating the WeGo database about Cloud Otto's current location and the app is reciprocally receiving information updated information about how to reach his "chosen" destination and various services he could avail himself of along the way. The crucial point here is that there is plenty of mutual information flow, but it is just now relevant to the sorts of cognitive vigilance a epistemically careful agent should want to be able to access.

27 For a lucid related discussion of transparency in relation to the mechanisms of consciousness and human evolution the reader is referred to (Metzinger, 2004, p. 61).

28 This point relates to some recent discussion over epistemic agency. Palermos notes that "in order for a process to be a candidate for inclusion to the agent's conscientious cognitive character, we noted it will probably have to be neither strange nor fleeting. (1) will have to be normal so that the agent won't reject it when conscientious and (2) will have to be a disposition or a habit of the agent." (Palermos, 2014, p. 1943). For Cloud-Otto it seems that these sorts of fleeting cognitive episodes are occasioned by his interaction with the We-Go app. Any standing believes about his destination for the day are at the mercy of the crowd of users of WeGo.

29 There is a relevant discussion of *procedural transparency* and *representational transparency* in Heersmink's (2015a) paper on the *Dimensions of integration in embedded and extended cognitive systems*. I have also discussed some related issues in (Clowes, 2013)

30 See Andrada (2019) for a detailed discussion on whether and how this problem might be resolved.

31 See (Andrada, Forthcoming) for a framework that analyses the interplay between organic capabilities, technologies and the sociocultural environment in which the interaction takes place.

32 Albeit this question is not uncontroversial; see discussion in (Clowes, 2017; Michaelian, 2012).

33 I am using the term bias here in a value-neutral way to mean something that effects or changes the character of a given cognitive process. This could be for the better, worse or be not easily assessed along a single evaluative dimension.

34 A good account of some of the way that AI technologies are being used to shape human decision architecture see (Russell, 2019).

35 In personal communication, Paul Smart points out that even the designers may not be aware of how some contemporary complex systems work, hence the interest in *Explicable Artificial Intelligence.*

36 There are somewhat contradictory trends at work in the development of many contemporary ICT systems. On the one hand there is a growing trend toward *self-tracking*, *i.e.*, the development of the quantified-self movement focused upon increasing "self-knowledge" in ways that depend upon the visualization of the data being generated by a variety of mobile devices as well as our interactions with a variety of Internet technologies (Bode & Kristensen; Lupton, 2016; Swan, 2013). The contrary trend

however is the growing tendency towards obscuring the computational bases–both in terms of data held on us by major online companies, and the actual algorithms that process this data–of computational systems run by the big companies such *Facebook, Google, Amazon*. Insofar as systems that might be part of our exoselves embody block reflective transparency, they will tend to undermine our regulative agency. The incorporation of very reflectively opaque cognitive systems tends to make us all look more like Cloud-Otto.