Horizons, PIOs, and Bad Faith

Abstract:

I begin by comparing the question of what constitutes continuity of Personal Identity Online (PIO), to the traditional question of whether personal identity is constituted by psychological or physical continuity, bringing out the compelling but, I aim to show, ultimately misleading reasons for thinking only psychological continuity has application to PIO. After introducing and defending J.J. Valberg's horizonal conception of consciousness, I show how it deepens our understanding of psychological and physical continuity accounts of personal identity, while revealing their shortcomings. I then argue that PIO must also be understood against the backdrop of the horizonal conception, that this undermines sharp dichotomies between online and offline identity, and that although online psychological continuity might become necessary for the preservation of our personal identities, we cannot become our PIOs. Finally, I argue that if PIO is understood solely in terms of psychological continuity, any increasing identification with our PIOs assumes the form of a paradigmatic project of bad faith: a technological reduction of our self-consciousness, rather than the enhancement it should be.

Keywords:

Personal Identity Online (PIO), personal identity, consciousness, Valberg, horizons, bad faith

I

Someone new to reflection on computer technology might be forgiven for thinking that the phenomenon of personal identity online (PIO) is quite unlike the ordinary personal identity of men and women, since only the latter requires flesh and blood. Long before there was any such phenomenon, however, philosophers were already arguing that personal identity has little to do with flesh and blood, thereby raising the prospect that PIO is simply a new side to a familiar old coin. I agree that a unified account of personal identity and PIO is indeed required, since there is no hermetically sealed virtual reality, no cyber-castle in the air that radically transcends offline mundaneness; the online and offline inevitably intermingle, as many philosophers have now realised (e.g. Dreyfus 2001; Ess 2011). However, I will be arguing, drawing on the powerful phenomenological insights of J.J. Valberg's 'horizonal' conception of consciousness, that this unified account must take full account of physical constraints. This paper will proceed as follows. Firstly I explain why developments in computer technology have seemed to lend weight to Lockean theories of personal identity, before, in section II, I present the *prima facie* case to think that PIO must be accounted for in Lockean terms. Lockean accounts are seriously flawed however, as I show in section III by introducing Valberg's theory of personal identity, albeit amended slightly as a physicalist theory. In section IV, I propose a further amendment, and then apply the resulting theory to the case of PIO. Finally, in section V, I suggest a possible element of bad faith underlying enthusiasm for pure Lockean theories, and related dreams of uploaded consciousness.

John Locke's psychological continuity theory of personal identity remains central to debates about how to specify the conditions in which a person continues to exist across time.[1] Locke's leading idea was that the criteria of identity we apply to judge the same man or woman has continued to exist are distinct from those we apply to judge the same person has continued to exist. The surprising upshot of this was that personal identity is less constrained by our biological natures than might have been thought, and at the present time, when computer technology is, in some sense at least, allowing us to construct our own identities online, Locke’s theory seems more relevant than ever. The question of its actual relevance, however, needs to be treated with considerable circumspection, as we shall see.

Locke distinguished the criterion of identity for a ‘mass of matter’, from that used to judge the continued existence of organic things, including human beings, which depends on the organisation of their matter. The criterion for personal identity, however, was something else again, and he sought to establish this with his thought experiment of the prince and the cobbler (1700 / 1979: 340). Thus Locke imagines a prince waking up one morning to find himself in the body of a cobbler, a body which now has all the memories and other psychological traits of the prince; we are invited to share Locke’s intuition that the prince is still the same person, even though he is no longer the same man. With the benefit of contemporary functionalist theories of mind, it is easy to fill in some science-fictional details; maybe the cobbler’s brain was erased, as we might erase the hard-drive of a computer, and subsequently reprogrammed with the psychology of the prince. Many philosophers were persuaded by this thought experiment long before the metaphor of the mind as software running on the hardware of the brain was available, but now that this way of thinking comes so naturally to us, Locke’s reasoning seems even more compelling.

As Locke puts it, it is ‘being the same consciousness that makes a Man be himself to himself, *personal Identity* depends on that only’ (II.27.§10). Thus sameness of consciousness is prioritised over physical continuity, and part of the reason why this came to seem increasingly relevant as the twentieth century progressed is that accounts of mind started to become increasingly detached from physical embodiment. This trajectory began with the functionalist reaction against the 1950s identity theory of U.T. Place (Place 1956), which had argued that mental states were physical states of the brain. The main problem with this theory, according to Putnam, was that it linked mentality too closely to human physiology, thereby making it unlikely that animals or extra-terrestrials could share our mental states, and effectively ruling out machine consciousness (Putnam 1967). In short, the identity theory was ‘chauvinistic’ (Block 1978), and functionalism sought to rectify this by construing mental states more abstractly, so they could be realised by other biological species or machines. Another significant decoupling of mind from brain came about with externalist theories of content, which had the consequence that the environment in which we think is partially determinative of what we think (Putnam 1970; Burge 1979). And more recently, extended mind theories have argued that cognition is a process that extends beyond the physical boundaries of the organism, thereby seeking to further break down the ‘hegemony of skin and skull’ (Clark and Chalmers 1998).

It is surely no coincidence that these moves away from reductionist accounts of mental states, have taken place during a time when computer science has been rapidly developing, and artificial intelligence research has been beginning to make headway. With the concept of mind seeming less constrained by the particularities of the human anatomy, then, and during a period when people were starting to routinely talk about what computers ‘recognise’, ‘know’ and ‘remember’, it is hardly surprising that Locke’s theory should have enjoyed a renaissance, since it held that personal identity is not determined by organic, human identity, but is rather a matter of psychological features that could be thought of as a programme, one that happens to be maintained by a human brain, but might also be maintained by a computer.

The most influential neo-Lockean theory is that of Derek Parfit, who motivates many of his ideas with the need to accommodate new technologies. Parfit's principal departure from Locke is that he thinks the concept of personal identity is overly restrictive. Thus in an imagined case of brain bisection, in which one half of a person's brain is placed into one human being, and the other into another, Parfit argues that regardless of the fact that neither of the resulting persons can be identified with the original, since this would violate the logic of identity, the original person has survived. This is because, assuming that both brain-halves maintain the psychological functioning of the original whole, the original person will be psychologically continuous with both new people. Given that the person prior to the operation does not seem to be facing death, Parfit concludes that psychological continuity is 'what matters' (Parfit 1984: 245& ff.); to this extent his view is thoroughly Lockean. Nevertheless Parfit wants to replace personal identity with the more flexible concept of survival, which is not an exclusively one-one relation, and which admits of degrees, such that a person may survive to a greater or lesser extent depending on the degree of psychological continuity maintained.

Our only concern for our individual futures, then, should be for the survival of our psychological traits, and Parfit thinks that belief in an 'extra fact' of personal identity, required for the preservation of an enduring self, is irrational. Parfit memorably illustrates this with his 'teletransportation' thought experiment, in which we are to imagine a machine that scans the physical state of a human body before destroying it, and then transfers the information to another machine which creates an exact replica. Since almost complete psychological continuity is preserved in this procedure, Parfit regards it as a way of travelling from the location of the first machine to that of second, and our strong intuition that the person using the machine has simply been killed and replaced by a replica is dismissed as irrational; we may prefer to keep the same brain and body, but this is just a matter of sentiment, like wishing to ‘keep the same wedding ring, rather than a new ring that is exactly similar’ (ibid.: 286). Parfit, then, is out to radically change the way we think about ourselves: we should care only about the survival of our psychological traits, for their own sake and not for the sake of an illusory enduring self, and thus we should not care whether the vehicle for this survival is the continued existence of our bodies, a teletransported replica, friends and family who we have influenced, or, of course, a machine.

Parfit regards this proposal as liberating:

> My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared. (Parfit 1984: 281)

Rather than viewing death as the final cessation of the conscious self, then, he suggests that we come to view it as simply the discontinuation of certain strands of direct psychological continuity, while other strands continue through other vehicles, such as people who remember and have been influenced by us. This view consoles Parfit by making death seem ‘less bad’, and provides a connection to others that removes the walls of his ‘glass tunnel’; he aligns it with the Buddhist doctrine of the unreality of self (ibid.: 502-3; see also Schopenhauer 1844 / 1969: 378), which also holds that liberation from egocentrism and consolation about the prospect of death are to be achieved by overcoming the illusion of an enduring self (see, e.g., Lui 2006: 209-247).

The Lockean conception of a person, then, leads us far from any notion of a physically-embodied, enduring self, and this has obvious resonances in a time in which virtual psychological continuities are maintained whose connections to physical reality are thoroughly transient. Moreover it might seem, as it does to Parfit, that distancing our self-conceptions from embodiment is liberating, and that consequently technological developments that aid in this process should be enthusiastically embraced. There are, however, dissenting voices. Thus Bernard Williams (1970) cast doubt on the kind of 'body-swapping' thought-experiment invoked to motivate Lockean views, arguing that acquiring the psychological traits of another person is not obviously transferral of personal identity, while Eric Olson (1999) has argued that personal identity is best understood without invoking psychological considerations at all; we are human organisms, and a human organism can survive the destruction of its psychological traits. In debates about offline personal identity, then, there is a standoff between psychological and physical continuity theorists.

II

Predictions abound to the effect that '[s]omeday "virtual-world" identities will be just as important as "real" identities – just as "e-commerce" has become indistinguishable from "commerce"' (Crawford 2006: 198). And it is not hard to see why, given the increasing amount of time and effort that many people put into fine-tuning their online presence; the kind of profile-management that was once only the concern of public figures is now becoming everybody's business. Moreover MMOGs (massively multiplayer online games) are an exponentially expanding phenomenon, with tens of millions regularly logging on around the world to interact with each other on a virtual rather than physical plane; people can now spend large proportions of their waking lives immersed in a ‘second life’, leading virtual lives more interesting than their monitor-bound ‘first life’, and identifying more with their avatar than their physical body. In short, PIOs are becoming increasingly important.

To understand the degree of control over identity the internet provides, we must distinguish two broad types of PIO, namely representative and non-representative. Thus a representative PIO is one which represents an offline person; a personal website, which provides biographical and other information about a person, contributes to a representative PIO for that person. These PIOs are factually constrained, to a greater or lesser degree depending on the kind of online activity engaged in, but in many cases leave plenty of creative scope for selective emphasis and interpretation. Non-representative PIOs, by contrast, are not supposed to represent an offline person, as for example when somebody creates a character in a MMOG. Although these two categories mark a significant distinction, however, there is considerable overlap, since a PIO may purport to be representative when it is not, or is only partially so, and ostensibly non-representative PIOs may in fact represent features of the offline person or persons maintaining them. The internet blurs the boundaries between fabrication and representation, much in the same way that fiction does.[2]

Suppose we want to give a theoretical account of PIO by providing the criteria of identity for a PIO to endure over time. Now in the case of regular personal identity, there are considerations favouring both psychological and physical continuity accounts, which philosophers have struggled to adjudicate. Thus, for example, Lockeans seem right that psychological continuity is all that matters to us, which can be seen from the fact that we tend to regard entering a state of complete and irreversible retrograde amnesia as equivalent to the death, even though the human organism remains alive. And yet there also seems an obvious sense in which we are our physical bodies, making it hard to believe that anyone could survive a complete loss of physical continuity; that is why most people baulk at the idea of teletransportation. However if we put personal identity to one side and turn our attention to PIO, then it might seem, initially at least, that the balance is thereby tipped firmed in favour of psychological continuity.[3]

A physical continuity account seems out of the question, because PIOs are constructed and maintained on physically disparate hardware vehicles, accessed using ISPs and routers throughout the world. Moreover, it seems likely that the ways in which the internet is physically maintained will eventually change beyond all recognition, and so even if particular physical continuities could be isolated now, they might later cease to exist; given that it is hard to see any reason why PIO continuity should not survive such changes, it seems these two kinds of continuity cannot be the same. An alternative to appealing to physical continuity of the PIO-vehicle would be to appeal to the physical continuity of the human being maintaining the PIO. But the problem with this is that any number of different human beings, or even machines, could contribute to the maintenance of a PIO. Could we instead appeal to the physical continuity of the human being represented by the PIO? Well, this would immediately exclude non-representative PIOs, and the proposal does not even work for representative PIO, because it seems clearly possible for the human represented to maintain its physical continuity while the PIO underwent such radical changes that its identity would be completely lost; online identity may be a matter of degree, just as Parfit thinks offline identity is, but there must still be some cut-off point, otherwise there could be no criteria of identity for PIO whatsoever, however vague or ephemeral.

Rather, than physical continuity, then, it is tempting, although as we shall see in the next section, ultimately misguided, to conclude that judgments of PIO sameness are based on broadly ‘psychological’ continuity: the PIO must simply manifest cognisance of previous online activity and display regular behavioural traits. This continuity may be a more or less accurate representation of the psychological continuity of the person maintaining the PIO, or simply a creation. Now test cases might arise in which a PIO forgets its past and starts acting out of character, which might make us wonder exactly how much psychological continuity is required. For example, if PIO-2 denies responsibility for the past online activity of PIO-1, but displays all of the distinctive character traits of PIO-1, might we be justified in judging that PIO-1 and PIO-2 are the same? Or what if PIO-2 lays claim to the history of PIO-1, but displays none of the character traits of PIO-1? It seems clear that we lack clear-cut criteria to decide such matters algorithmically, perhaps as a matter of principle, if such judgements involve an ineliminable element of *phronesis* (cf. Ess and Thorseth 2011: xxiii), or if the identity-conditions of PIOs are inherently vague; alternatively it might simply be that there are too many possible variables to make the formulation of strict criteria practicable. But whatever the reason, this lack of strict criteria for reidentification does not count against psychological continuity accounts *per se*, since on Parfit’s account, for instance, survival is not an all-or-nothing affair but rather a matter of degree, and so we may judge that a PIO has survived to a greater or lesser extent without having to answer elusive questions about sameness.

Now supposing we were to conclude that PIO continuity is indeed best accounted for in terms of psychological continuity, we might nevertheless still be of the opinion that this is of no great consequence, on the grounds that PIO is not real personal identity: it is just a public profile of a person in the case of representative PIOs, or a fictional character in the case of non-representative PIOs. Thus if I radically change my online profile, or terminate my old MMOG character and introduce a new one, then in both cases we might want to describe this as the end of one PIO and the beginning of another. But even if it suits us to talk this way, this situation still seems in no way comparable to the real loss of identity that can occur as the result of serious brain damage, or when a human being dies, for example.

This conclusion may be premature, however. Suppose a pioneer emerges, call him Jack, who spends almost his entire waking life online; he never interacts with anybody except online, and his offline life simply consists in performing, zombie-like, the series of routines required to keep his body ticking over. Jack's whole psychology, we may suppose, is both directed and manifested online. Thus he only remembers events in his online life, since one day is just like the next as far as his offline routines are concerned, and he exhibits no psychological traits offline. Given that the only psychological continuity maintained throughout Jack's life, then, is his PIO, it seems reasonable to conclude that if personal identity is a matter of psychological continuity, as Lockeans believe, then Jack is his PIO. After all, Jack's self-concern is for his PIO only, and likewise it is only Jack's PIO that his friends know or care about; arguably there is nothing else distinctive about Jack *for* anyone to care about.

Now suppose that further technological advances allow Jack's descendants to short-circuit the clumsy interface with online activity provided by the human body: brains are now hardwired directly into the internet. Biological brains still have to die, however, thus introducing what might well seem to Jack's descendants to be an arbitrary termination to PIO, but using artificial neural prostheses, let us suppose that it becomes possible to replace the brain with non-biological material performing the same functions; the neural prostheses can maintain a PIO indefinitely. It may still seem to us that the hardware is the real repository of personal identity, whether this is a biological brain, or a functionally equivalent system, but perhaps this is just a manifestation of the irrational belief in an 'extra fact' of personal identity that Parfit warned us against. Perhaps our only concern should be with continuity of PIO, and we should be indifferent to the continuity of the hardware required to maintain it.

If we take Lockean accounts of personal identity seriously, then, the possibility of transferring our psychology online seems to have major philosophical implications. That we seem to be experiencing the beginnings of this transferal might even be interpreted, Hegelian-style, as part of an inexorable progression towards 'leaving the meat behind' to 'exist as pure data or uploaded consciousness' (Bell et al. 2004: 8); perhaps technology is helping us to become self-conscious about our true natures as 'centers of narrative gravity' (Dennett 1992). And maybe such developments are to be welcomed, since displacing our self-conceptions from the physical bodies we were born with, to an online psychological continuity we freely invent, might be considered both equalising and liberating. However, although there is much to be said for this kind of optimism, there is another account of personal identity which should first be considered, since it puts the above reflections into a new perspective. This is J.J. Valberg's 'horizonal' account of personal identity.

III

Valberg's account of personal identity, which reveals the shortcomings of Lockean theories, is a consequence of his conception of consciousness as the horizon of experience. An instructive way to think about the horizonal conception is that it is the result of systematically thinking through what it would mean for our perception of the world to be direct, rather than mediated by internal ideas. Much of twentieth century philosophy of mind can be seen as a reaction to the seventeenth century conception of consciousness as a phenomenal array of subjective ideas providing indirect epistemic access to the objective world. This 'phenomenal' conception, as Valberg calls it, is both epistemologically and ontologically problematic. Epistemologically, it apparently rules out any possibility of determining whether our ideas correspond to reality; as Locke himself acknowledged, 'the having the idea of anything in our mind no more proves the existence of that thing, than the picture of a man evidences his being in the world' (Locke 1700 / 1979: 630). And ontologically, the problem with subjective ideas is that they seem to have no place within the physical world described by modern science. In order to avoid these problems, philosophers have tried to undermine arguments purporting to show that perception must be indirect, and have pioneered theories of mind, such as behaviourism and the identity theory, which reject subjective ideas. Valberg's contribution is to show that we have an alternative, 'implicit' conception of consciousness, which 'even if we never articulate it (…) is with us all the time' (Valberg 1992: 124), and which is more readily able to accommodate direct perception.[4]

Valberg motivates the horizonal conception, which as we shall see provides a combined phenomenological / causal account of personal identity, with his 'argument from internality'

(Valberg 2007: 27 & ff.), which begins in reflections on the nature of dream-scepticism. When we enter into traditional Cartesian meditations and ask ourselves whether this is all just a dream, what exactly, Valberg wonders, do we mean by ‘this’? Evidently not some particular object we might be focusing on while asking this question, such as a tree. Rather, we are asking whether the tree, along with the totality of space and time it belongs to, exists in a dream or reality. Thus if this is a dream, the totality of space and time the tree belongs to exists within a dream, and if this is reality, the totality of space and time the tree belongs to exists within reality. What we mean by ‘this’, then, is not something within the totality of space and time, or even the totality itself. Rather, we mean the experiential horizon within which the totality exists: we are asking whether it is the experiential horizon of a dream or of reality. If it is the horizon of a dream, it may be ‘displaced’ when we wake up by the ‘wider’ horizon in which actual space and time exists. If ‘this’ is the horizon of reality, however, then there is no wider horizon to displace it; reality itself exists within this horizon.

Now if anything we can pick out within the horizon of consciousness is part of the world, whether a dream-world or the real world, what is consciousness itself? We cannot be conceiving it as part of the world, because the world appears within consciousness. Thus consciousness must rather be ‘“that within which” the world is present’ (Valberg 1992: 125; Valberg 2007: 97). This means that there is a sense in which we conceive consciousness as not a part of the world, but rather a nothingness, and Valberg finds clear precedents for this view in Kant’s conception of the transcendental self (Valberg 2007: 13-4 & 400-7), Wittgenstein’s conception of the metaphysical subject as not part of the world but its ‘limit’ (Valberg 1992: 124-5), and most transparently, Sartre’s description of consciousness as a ‘nothingness’ (Valberg 2007: 14). The reason consciousness is described as a ‘nothingness’, is that it is nothing apart from the world’s presence within it, and hence nothing in itself;

consciousness is not a part of the independently existing totality of space and time, but rather the horizon within which this totality exists as present. As such, consciousness can be thought of as a context which qualifies independently existing objects as present, rather as society provides a context which qualifies rocks as milestones (Valberg 1992: 122).

In place of the early modern conception of experience as the direct observation of subjective appearances of a consciousness-transcendent reality, then, the horizonal conception allows us to think of experience as the direct observation of the world within the context provided by a spatiotemporally transcendent horizon of consciousness. Anything I can pick out within consciousness is a part of the world, rather than a feature of consciousness itself, which means that the world is revealed rather than hidden by consciousness. What is important about this conception of consciousness from our present perspective is that as soon as we make it explicit in our thinking, it provides a compelling account of personal identity which offers an alternative to physical and psychological accounts, while also deepening our understanding of them.

Valberg agrees with the Lockean psychological continuity account that personal identity requires sameness of consciousness; the difference is in how this sameness is conceived. On the horizonal account, I continue to have the same consciousness so long as the world can exist within my horizon, and hence can appear or be present to me; this is compatible with there being periods of unconsciousness when the world stops appearing (Valberg 2007: 178-9). To understand the conditions under which the world is able to continue appearing within my horizon requires a combination of first- and third-person reflection: I must reflect on how the future might develop within my horizon, and I must reflect on the causal conditions required to sustain my horizon. Such reflection yields both a 'horizonal' and 'positional' component to personal identity.

The horizonal component is straightforward: I exist only so long as my horizon exists. If I reflect in the first-person on what is required for my life to continue, then, I must reflect on experience temporally unfolding within my horizon. However, third-person reflections are also relevant, since we know as an empirical fact that the horizon of consciousness exists only so long as our physical bodies, and in particular, our brains, continue to exist (Valberg 2007: 222 & ff.) The explanation of this fact which I shall presuppose here is the physicalist one that consciousness *is* the brain: our conceptions of experiential horizons and biologically functioning brains are of course radically different, but to conceive is to represent (cf. Laurence and Margolis 1999), and we can represent the same things in radically different ways.[5] Given this identity, then, the prospect of the destruction of my brain is the prospect of the destruction of my horizon, and hence of 'once-and-for-all NOTHINGNESS' (ibid.: 229); Valberg capitalises the word to signify a complete lack of experiential presence. However, there is also another component to personal identity, since although my continued existence requires the continued existence of my horizon, there is no sense in which I *am* my horizon; my horizon is an experiential context for the world, and it would be a category mistake to identify myself with such a 'thing'. Rather, and here Valberg agrees with physical continuity accounts, I am a human being, and which human being I am is determined positionally: I am the one 'at the center of my horizon' (ibid.: 337), where 'centrality' is provided with a complex phenomenological analysis which distinguishes my willed body from other more peripheral objects within my horizon (ibid.: 286-320).

This integrated two-component account reveals the main shortcoming of psychological continuity accounts, such as Locke’s original account of personal identity and Parfit’s account of survival, which isolate psychological continuity from the causal conditions required for that continuity to take place within a unitary experiential horizon. Thus consider the ‘body-swapping’ case of the prince and the cobbler. If we think of psychological continuity as a temporal sequence of discrete occurrent phenomenal and cognitive states, along with the preservation of dispositional states such as memory and standing knowledge, then there seems no principled reason why a transfer of consciousness should not take place; the cobbler’s body must simply instantiate type-identical states to those once instantiated by the prince’s body, i.e. exactly the same kinds of states, together with others that appropriately continue the sequence. If, however, we stop thinking about the continuity of a sequence of states, and instead imagine ourselves into the first-person perspective of the prince prior to the alleged body-swapping, we find that it is not so clear how the prince is to supposed to realistically imagine his future experience panning out in such a way that, for example, his current view of his palace is instantaneously replaced with a view of the cobbler’s workshop. The fact of this experiential discontinuity, since experiences of travelling from the palace to the workshop are missing, tips us off to the presence of a causal abnormality.

Then when we consider the situation from a third-person perspective, we find that there is nothing to make causal sense of the idea that the prince’s horizon has been transferred to the cobber’s body. With enough science-fiction license, we can make sense of the physical constitution of the cobbler being transformed so that it now instantiates type-identical psychological states to the prince. But the continuity of the prince’s horizon depends on his own particular brain, and not the distinct brain of the cobbler, and so no matter what physical transformations the cobbler’s brain undergoes, it could never sustain a particular horizon that is not its own. This is a simple consequence of the fact that, if we combine the horizonal conception with physicalism as I am suggesting, the prince's horizon *is* his brain. As such, although the cobbler-human-being may have become psychologically indistinguishable from the prince, in the sense that he thinks and acts exactly as the prince would, that particular human being cannot be the prince.

The same considerations apply in Parfit's teletransportation case; if we take up the first-person perspective of the teletransporter, it is hard to realistically imagine that my experiences of one place might instantaneously be replaced by experiences of another, without any intervening experiences of travel. And then when we reflect on the situation causally, we remember that teletransporting will destroy the particular brain which supports my horizon; as Valberg puts it, '[i]nsofar as the prospect that I face includes the destruction of my brain, it includes the prospect of absolute and final NOTHINGNESS' (ibid.: 443). Parfit's mistake is that by thinking of consciousness as a sequence of states, in accordance with the phenomenal conception, he imagines certain psychological states pre-teletransportation, then others post-teletransportation, and finds an almost perfect continuity across the procedure, barring what seems from this third-person perspective to be just an experiential glitch. However, when we remember that the moment I teletransport is the moment at which my brain is destroyed, and hence my particular horizon ceases to exist, we see that psychological continuity within my horizon, which is the only kind of psychological continuity relevant to my continued existence, would at this point cease; Parfit has made the mistake of 'jumping in imagination over that over which there is no jumping' (ibid.: 443-4).

Now not all Lockean accounts isolate psychological continuity from causal considerations. For example, Parfit's 'Psychological Criterion' requires both psychological 'strong connectedness' and that this connectedness 'has the right kind of cause' (Parfit 1984: 207). 'Strong connectedness' is supposed to capture the idea of there being a sufficient number of direct psychological links for personal identity to be preserved; Parfit recognises that 'we cannot plausibly define precisely what counts as enough', but nominally specifies the requirement as '*at least half* the number [of direct psychological connections] that hold, over every day, in the lives of nearly every actual person' (ibid.: 206). This condition, then, is evidently met in the body-swapping and teletransportation examples; the momentary experiential discontinuity would be massively outweighed by the otherwise perfect continuity presupposed by these cases. In addition to 'strong connectedness', there is also the causal condition, which varies in three different versions of the psychological criterion. According to the 'wide' version, the cause must be reliable, and there is no principled reason why teletransportation and body-swapping should not be, while according to the 'widest', which Parfit himself adopts for his own account of survival, no causal restrictions are made at all. The 'narrow' version, however, requires that psychological continuity 'have one of several normal causes', thereby ruling out 'abnormal interference' (ibid.: 207). As such, a 'narrow' psychological continuity has the resources, just like the physicalist horizonal account, to deny that personal identity is preserved in the body-swapping and teletransportation cases.

What is unique about the horizonal account, however, is the explanation it provides of why causal restrictions are necessary, and of the sources of the appeal, as well as the shortcomings, of both psychological and physical continuity accounts. According to this explanation, 'pure' psychological continuity accounts, which impose no causal restrictions, are right to the extent that my continued existence requires only the continued existence of my consciousness. However by misconceiving consciousness, such accounts overlook the fact that this continuity requires the continued existence of a particular brain; ‘narrow’ theories are able to rectify this, but without a unified phenomenological / causal account of the kind provided by the horizonal account, a restriction upon the causes of psychological continuity seems *ad hoc*.[6] Physical continuity accounts, on the other hand, do have a principled reason to require the continued existence of a particular brain, and are right to hold that a person is a just particular human organism; the horizonal view also endorses this common-sense view. However by neglecting the importance of consciousness, such accounts overlook the fact that it is only in the positional sense that I am a particular human organism. This seems to leave open the possibility, exploited by the Lockean thought-experiments, that a new body could come to occupy the centre of my horizon.

Whether this really is a possibility depends on whether horizonal continuity can be preserved from one body to another; there is a sense in which Valberg thinks it can, but only if the living brain were physically removed from one body and transplanted into another:

> But what if we suppose that (…) the brain of the human being who occupies the subject position within my horizon, is transplanted to the body of another human being? I would find myself in his body. This may not be causally possible, but here, it seems [we have] described a case that is not just experientially but metaphysically possible. (ibid.: 451)

It is experientially possible in the sense that I can coherently imagine my future panning out this way: my perceptions would cease when my brain was removed, leaving me with only thoughts and mental images, but I would perceive again through the new body once the transplant was complete. This case is crucially different from the teletransportation case, in which my brain is destroyed and subsequently replicated, since in the transplantation case, my brain, which accounts for horizonal continuity, continues to exist. The transplantation case is also metaphysically possible, since it involves no conflict with metaphysical principles such as the law of sufficient reason. But nevertheless, for all we know, it may be ‘causal nonsense’ (ibid.: 450); it may not be possible to do this kind of transplant given the actual physical laws that govern the world. Either way, the fact that we can at least imagine such cases provides insight into the combined horizonal and positional components of personal identity, while also explaining the appeal of psychological continuity theories.

IV

We saw in section II that psychological continuity seems, on the face of it, considerably more relevant to PIO than physical continuity. If we now ask how the horizonal account applies to PIO, it might seem, again on the face of it, just as inapplicable as physical continuity. Moreover, since the horizonal account has the consequence that personal identity cannot be accounted for in terms of psychological continuity, we might draw the additional conclusion that personal identity and PIO are thoroughly disparate phenomena. As we shall see, however, the horizonal account in fact provides the resources to provide a unified account of personal identity and PIO, and one which relies integrally on physical continuity.

The reason the horizonal account might seem inapplicable to PIO, is that it requires continuity of a unitary horizon, and yet on the face of it, PIO does not, since a PIO could be maintained by different people or machines. As such, it seems that a PIO could exist without existing in any one particular horizon, and perhaps even any horizons at all.[7] Horizonal continuity is of course relevant to representative PIOs, given that these PIOs represent offline personal identities; horizonal continuity would thus be a determining factor of the representational accuracy of this kind of PIO. However, it seems, according to this line of reasoning, that although PIO can represent offline personal identity, the two are otherwise unconnected: PIO requires psychological continuity, in some cases with the additional requirement that the PIO represent an offline person, whereas offline personal identity requires the continuation of an experiential horizon, which in turn requires the continuation of a functioning biological brain. Lockean theories which purported to account for personal identity, then, apparently fail in the task they were designed for, but come into their own online.

Matters are not so simple, however, as becomes apparent when we reflect on the fact that psychological continuity, in the Lockean sense of continuity of memories and other psychological traits, has no essential role to play within Valberg's account.[8] However, it is one thing to reject the claim that psychological continuity is all that is required for personal identity, and quite another to exclude it from an account of personal identity altogether. After all, we have the strong intuition that losing all of our psychological traits through irreversible brain damage amounts to the end of our personal identity. Since the prospect of this kind of brain damage seems to be the prospect of a person ceasing to exist, although not a human being or their horizon ceasing to exist, it is hard to see how personal identity could have nothing to do with psychological continuity in this non-horizonal sense.

Although Valberg seems right that personal identity requires horizonal continuity, then, it also seems to be the case that I would not survive the loss of all my memories and other psychological traits even if my horizon did continue to exist: the prospect of losing all my psychology is distinct from the prospect of 'absolute and final NOTHINGNESS', but both seem sufficient to put an end to me. Now this conclusion might be rejected on the grounds that a person is simply a human being, and since human beings can continue to exist even after their psychological continuity is lost, so can persons; this argumentative strategy is often central to physical continuity accounts (cf. Olson 1999). However, Valberg does not claim that being a particular human being is essential to personal identity; in the qualified concession to the Lockean tradition we saw earlier, he allows the metaphysical possibility of a brain transplant in which a person continues to exist while the human being at the centre of their horizon is replaced. This conceptual distance between being a person and being a human being allows the horizonal account to be supplemented with Lockean psychological criteria.

To accommodate the necessity of psychological continuity, then, we can simply say that I am the human being at the centre of my horizon, and psychological continuity within that horizon is necessary for the human being that I am to remain the same person. Thus I am both a human being and a person: I could remain the same person but cease to be the same human being if my brain were transplanted into another human being, and I could remain the same human being but cease to be the same person if I were to enter a state of complete and irreversible retrograde amnesia. Lockean psychological continuity is a necessary condition for the preservation of personal identity, then, but not a sufficient one, since horizonal continuity is also required; that is why personal identity is not preserved in teletransportation, since the psychological states of the replica would not exist in the same horizon as the states of the person who provided the blueprint. Psychological continuity provides a criterion of personal identity only within a horizon, then, for otherwise it cannot count as *my* psychological continuity, and this requirement can be seen as another way of putting Kant's fundamental insight into consciousness, namely that 'it must be possible for the "I think" to accompany all my representations' (Kant 1787 / 1933: B131; cf. Ess and Thorseth 2011: xx). In the *Paralogisms*, Kant made it quite clear that this purely 'formal' identity, which Valberg identifies with horizonal continuity (Valberg 2007: 400-7), is not to be confused with either sameness of substance or psychological continuity.[9]

Once we understand personal identity this way, PIO no longer seems a disparate phenomenon. In the case of non-representative PIOs, a pure Lockean account remains the best option, since the identity of the PIO, like any fictional character, is independent of the identity of whomever or whatever maintains it.[10] In the case of representative PIOs, however, horizonal continuity is also crucial, since it is the determining factor in whether a PIO counts as *my* PIO, just as it is the determining factor in whether offline psychological continuity is mine, as opposed to, for example, the type-identical psychological continuity my teletransported replica possesses. Thus if I express my psychological traits online, and the experiences associated with doing so take place within a unitary horizon, the PIO is mine. Other conscious subjects, or even machines, could accurately represent the facts about my life, and in principle could perform exactly the same online interactions I would, but without a background of horizonal continuity, this could not be constitutive of my PIO. Thus even if others judged, on the basis of psychological continuity, that my PIO had outlived my physical death, the passing of my horizon would mean that it was no longer the same PIO.

The horizonal account reveals the substantive sense in which we may be said to possess an identity online, then, but the intriguing question remains of whether I could become my PIO, as the pure Lockean account seemed to imply. Well, if offline psychological continuity can provide a necessary condition for me to remain the same person, there seems no reason in principle why online psychological continuity should not do the same were my engagement with the internet to become so thorough that I could only express my psychological traits online; this was the situation for Jack. However even if continuity of PIO did become necessary to our personal identities, we would still remain human beings. Thus when we imagine the first-person experience of the person represented by the PIO, all there is to imagine is the experiences of a human being sat at a console and performing the bodily actions required by his or her online activity (cf. Søraker 2011: 62-4). This person has a first-person perspective because he or she is a conscious human being, but the PIO is not a human being: it is an online expression of the psychology of the human being.

That is not quite the end of the story, however, because the positional component of the hozional account opens up the possibility of PIO becoming even more integral to what we are. For as we become more immersed in virtual reality, it seems both experientially and metaphysically possible that an avatar might come to occupy a position at the centre of my horizon, analogous to the position my physical body currently occupies. To do so it would need, at a minimum, to become the locus of my perceptual field and agency. But we already know that something very much like this is possible, because in dreams our brains generate a virtual world we can perceive and intentionally interact with from the perspective of a dream-body. If an avatar were at the centre of my horizon and my psychology was all expressed online, then, would I be my PIO? No, because just like a dream-world, this virtual-world could be displaced by the wider horizon of reality, which is the horizon that positionally determines what I am, namely a human being. And even if my brain was plugged directly online and I no longer had a body, the question of what I am would still be settled in the horizon of reality: I would be a brain.

V

If we fail to understood PIO against the backdrop of horizonal identity, and instead think of it in terms of pure Lockean psychological continuity, then in coming to identify more with our PIOs and less with our bodies, we risk losing sight of our nature as conscious human beings. Moreover, there is a clear and powerful motivation to do this, of a kind which most of the major world religions have addressed themselves to, since if we misconceive ourselves as a stream of online psychology, we thereby think of ourselves as something which need never die, since this kind of psychological continuity can be maintained by other people or by machines. By embracing this self-conception, then, we distance ourselves from the feature of the human predicament that people have always found the most disturbing, and which religions have always sought to console us about with doctrines of reincarnation and the afterlife, namely mortality. This is not to deny that there are solid reasons in favour of the Lockean view, of course, for Locke's essential point that our concepts of a person and of a human being are associated with distinct criteria was a powerful and far-reaching one. The claim is simply that there is a strong, religious motivation to bring oneself around to this kind of view, as Parfit’s expression of Buddhist consolation suggests: there is a potentially life-changing personal reward to be gained from reaching such conclusions. As such, any reasoning which seems to show that mortality is not a fundamental aspect of the human condition should be subjected to extra scrutiny, since we might be falling for a motivated self-misconception. And now that technology is severing the link between psychological continuity and our biological bodies, this temptation to misconceive ourselves, if that is what it is, has become easier than ever to embrace.

The existentialists had a name for self-misconceptions motivated by the desire to negate some aspect of our fundamental situation: bad faith. Thus in the best-known example of bad faith, Sartre describes a waiter acting out his role like an automaton, in an attempt to negate his freedom and hence responsibility for his actions. The reverse-side of this kind of bad faith, which is more relevant here, is when we overemphasise our freedom in an attempt to negate our ‘facticity’, that is, the unnegotiable facts about ourselves such as that we are embodied, have a personal history, and exist within a certain social environment. Sartre’s example of this kind of bad faith involves a woman who attempts to abdicate responsibility for the situation she is in by ‘disowning’ her body when the man she is dating takes her hand; she ‘leaves her hand there, but she does not notice (...) because it happens by chance that she is at this moment all intellect’ (Sartre 1943 / 1969: 55-6). It is this kind of facticity-denying bad faith that an over-identification with PIO threatens. Images of unhealthy people preening their pristine avatars provides an obvious manifestation of this, but there are many more subtle ways to negate your facticity online, such as by losing touch with all but your airbrushed online history, or by disassociating yourself from the particularities of your local community in favour of a more anonymous and ephemeral online culture.[11]

Bad faith is a form of self-deception which weakens our connection to reality. Thinking about PIO within the framework provided by the horizonal conception, however, helps us avoid over-emphasising the freedom provided by online activity to the detriment of our biological and socio-historical facticities. Computer technology provides us with a massive and unprecedented potential to expand and enhance our personal identities within a wider inter-personal arena than offline interactions could ever provide, but it is important that we do so self-consciously, and do not end up using the technology to suppress what we already know about ourselves. However although the horizonal account dampens down this tendency towards bad faith, by keeping our identities firmly rooted in our biological natures, it also, unlike pure physical continuity accounts, leaves room to show how PIO might become integral to our identities, which is important if, as now seems all but inevitable, we are going to be spending increasing amounts of our time expressing and developing our psychological characteristics online.

Word count: 8262

References:

Balkin, J. and Noveck, B. (eds.) (2006) *The State of Play: Laws, Games, and Virtual Worlds*, New York: New York University Press.

Bell, D., Loader, B., Pleace, N., and Shuler, D. (eds.) (2004) *Cyberculture: The Key Concepts*, London: Routledge.

Block, N. (1978) ‘Troubles with Functionalism’, in C. Wade Savage (ed.) *Minnesota Studies in the Philosophy of Science*, 9: 261-325.

Burge, T. (1979) ‘Individualism and the Mental’, *Midwest Studies in Philosophy*, 4: 73-122.

Clark, A. and Chalmers, D. (1998) ‘The Extended Mind’, *Analysis*, 58: 7-19.

Crawford, S. (2006) ‘Who’s in Charge of Who I Am? Identity and Law Online’, in J. Balkin and B. Noveck (eds.) (2006), pp. 198-216.

Dennett, D. (1992) ‘The Self as a Center of Narrative Gravity’, in F. Kessel, P. Cole and D. Johnson (eds.) *Self and Consciousness: Multiple Perspectives*, Hillsdale, NJ: Erlbaum.

Dreyfus, H. (2001) *On the Internet*, London: Routledge.

Ess, C. (2011) ‘Self, Community, and Ethics in Digital Mediatized Worlds’, in C. Ess and M. Thorseth (eds.) *Trust and Virtual Worlds: Contemporary Perspectives*, New York: Peter Lang.

Ess, C. and Thorseth, M. (2011) ‘Introduction’ in their (eds.) *Trust and Virtual Worlds: Contemporary Perspectives*, New York: Peter Lang.

Kant, I. (1787 / 1933) *Critique of Pure Reason*, trans. N. Kemp Smith, London: Macmillan.

Laurence, S. and Margolis, E. (1999) ‘Concepts and Cognitive Science’, in their (eds.) *Concepts: Core Readings*, Cambridge, MA: MIT Press.

Locke, J. (1700 / 1979) *An Essay Concerning Human Understanding*, P. Nidditch (ed.), Oxford: Clarendon Press.

Lui, J. (2006) *An Introduction to Chinese Philosophy: From Ancient Philosophy to Chinese Buddhism*, Oxford: Blackwell.

Olson, E. (1999) *The Human Animal: Personal Identity without Psychology*, Oxford: Oxford University Press.

Papineau, D. (2000) 'The Rise of Physicalism', in M. Stone and J. Wolff (eds.) *The Proper Ambition of Science*, London: Routledge.

Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.

Place, U.T. (1956) 'Is Consciousness a Brain Process?', *British Journal of Psychology*, 47: 44-50.

Putnam, H. (1967) 'The Nature of Mental States', in W. Capitan & D. Merrill (eds.) *Art, Mind, and Religion*, Pittsburgh University Press. Reprinted in Putnam 1975: 429–40.

Putnam, H. (1970) 'Is Semantics Possible?', in *Metaphilosophy*, 1: 187- 201. Reprinted in Putnam 1975: 139-51.

Putnam, H. (1975) *Mind, Language, and Reality*: *Philosophical Papers Volume 2*, Cambridge: Cambridge University Press.

Sartre, J-P (1943 / 1969) *Being and Nothingness*, trans. H. Barnes, London: Routledge.

Schechtman, M. (1996) *The Constitution of Selves*, Ithaca, NY: Cornell University Press.

Schopenhauer, A. (1844 / 1969) *The World as Will and Representation*, vol. 1, trans. E.F.J. Payne, Toronto: Dover Publications.

Shoemaker, S. (1984) ‘A Materialist’s Account’, in S. Shoemaker and R. Swinburne (eds.) *Personal Identity*, Oxford: Blackwell.

Søraker, J. (2011) ‘Virtual Entities, Environments, Worlds and Reality: Suggested Definitions and Taxonomy’, in C. Ess and M. Thorseth (eds.) *Trust and Virtual Worlds: Contemporary Perspectives*, New York: Peter Lang.

Spaight, T. (2006) ‘Who Killed Miss Norway?’, in J. Balkin and B. Noveck (eds.) (2006), pp. 189-197.

Valberg, J.J. (1992) *The Puzzle of Experience*, Oxford: Clarendon Press.

Valberg, J.J. (2007) *Dream, Death, and the Self*, Princeton, NJ: Princeton University Press.

Williams, B. (1970) ‘The Self and the Future’**,** *Philosophical Review,* 79: 161-180.

Wittgenstein, L. (1958) *The Blue and Brown Books*, Oxford: Basil Blackwell.

[1] This is probably the most common issue debated under the heading of 'personal identity', although there are many other related issues which fall outside the scope of this paper; as Wittgenstein noted, 'the term "personality" hasn't got one legitimate heir only' (Wittgenstein 1958: 62; see also Schechtman 2007: 1-2).

[2] Tracy Spaight's 'Who Killed Miss Norway?' provides an interesting case-study to illustrate this point (Spaight 2006); although a large online community was convinced a former 'Miss Norway' had been killed, the rather more mundane reality was (probably) that a Norwegian man decided to withdraw his non-representative PIO from a MMOG.

[3] That offline and online identity might well appear distinct, on the grounds that physical continuity is apparently less relevant to PIO, suggests a hard dichotomy between the online and offline. However a consensus has now developed that such a dichotomy, routinely presupposed in the 1990s, is not tenable; Charles Ess, for instance, sees it as a recent incarnation of Cartesianism, which was undermined by computer-mediated communication research that 'extensively and intensively documented the multiple ways in which the offline and online more and more seamlessly interweave with one another rather than stand in sharp, 1990s-style opposition' (Ess 2011: 23). One of my main aims here is to build on this consensus, by showing that physical continuity has more relevance to PIO than first meets the eye, and that online and offline identity are fundamentally intertwined.

[4] Valberg does not motivate the horizonal conception as I am doing here, and never seeks to 'deny the validity' of the phenomenal conception, on the grounds that we may legitimately talk of conscious *states* (Valberg 2007: 99). But if these states present subjective phenomenal properties rather than the world, it is far from clear that the two conceptions are compatible. Moreover, a commitment to the existence of conscious states does not entail that these states are phenomenal; the original 1950s identity theory denies that they are, for instance (cf. Place 1956: 49).

[5] Valberg himself does not make this kind of physicalist claim, and says only that the brain causally maintains the horizon of consciousness, although he does argue (in an unpublished manuscript, *Reflections on the Nature of Mind*) that the horizonal conception is *compatible* with physicalism. In common with many contemporary philosophers of mind, however, I take the minimal claim of token-identity, which is common currency between the various kinds of reductive and non-reductive physicalist theories, to be the simplest and least problematic explanation of the empirical fact that each person's state of consciousness systematically covaries with the state of their brain. The dualist alternative of accounting for horizonal continuity in terms of immaterial substance would have to address the substantial evidence for the causal completeness of physics; see Papineau 2000.

[6] This is not to deny that there could be other principled reasons for imposing causal restrictions. Shoemaker, for instance, defends a narrow psychological account which motivates its causal restrictions with a functionalist account of mental states (Shoemaker 1984: 92-101).

[7] This latter claim might be denied on the grounds that a PIO only exists if interpreted as a PIO, and that this interpretation can only take place within a horizon; whether this is plausible turns on the wider issue of whether information is consciousness-dependent or -independent.

[8] Valberg provides an insightful critique of Locke's conception of continuity of consciousness (Valberg 2007: 376-80), but neglects a crucial feature of this view, namely that it is the psychological content of consciousness that individuates a human being as a *person*.

[9] This latter point is clearest in the famous footnote to the third *Paralogism*, where Kant imagines a succession of substances passing on their psychological states from one to another, such that the 'last substance would then be conscious of all the states of the previously changed substances, as being its own states' (A364); this kind of psychological

continuity is thereby contrasted with the purely formal identity through time required by the transcendental unity of apperception.

[10] It is crucial to note here that non-representative PIO, as set out in section II, is the identity of a fictional entity, and not the identity of the person or persons contributing to that identity. The two might overlap: the same fictional character (in a MMOG, for instance) might be considered as a representative (if perhaps highly fictionalised) PIO for me, but also as a non-representative PIO, i.e. the character his- / her- / it-self. In the latter sense, I am evidently *not* my PIO even if I alone maintain it, any more than Charles Dickens *is* Oliver Twist, however much autobiographical content Dickens put into his character (if this is not immediately obvious, consider the fact that Dickens and Twist have different parents). It is only in the non-representative sense, then, that I am advocating a Lockean account, and this is due to the nature of fiction rather than anything specific to ICT; in the more interesting, representative sense of PIO, however, no sharp distinction between online and offline identity can be drawn.

[11] Escapism can of course be quite innocent: we can thoroughly absorb ourselves in a fantasy, and there is no reason, given the right circumstances, why someone should not, without fear of moral reproach, regard an online fantasy as the most important thing in their life. Escapism need not involve self-deception, however. But even when it does, the moral connotations of bad faith do not imply moral inexcusability; the fault may be very mild. My aim, much like Sartre's (I think), is not to condemn bad faith morally but intellectually, on the grounds that it reduces our self-consciousness: whether the moral consequences of this, all things considered, are good or bad, the consequences for our understanding are clearly bad.