The Philosophical Quarterly Vol. 00, No. 0
ISSN 0031-8094

2021
https://doi.org/10.1093/pq/pqab037

PERSISTENCE WITHOUT PERSONHOOD: A NEW MODEL

By Joseph Gottlieb

I am a person. But am I fundamentally and essentially a person? The animalist says no. So must the phenomenal continuity theorist, or so I'll argue. Even if, contra animalism, we cannot survive zombification, being a subject of experience is not sufficient for being a person, and phenomenal continuity is not sufficient for our survival as the same person. These observations point the way to a positive account of personhood, and provide further insight into the conditions under which literal survival preserves what matters.

Keywords: personal identity, phenomenal continuity, consciousness, access consciousness, vegetative state, forensic personhood.

I. FORENSIC PERSONHOOD

Let *person essentialism* be the following *de re* thesis: if *x* is a person, *x* cannot exist at any time *t* without being a person at *t*. So, since *we* are persons, person essentialism is a claim about what *we* are fundamentally. It says that we are fundamentally persons.

'Person' is a slippery term. Some use 'person' to refer to whatever we are most fundamentally, making person essentialism trivially true. However, the usage adopted here can be set out like this:

> **Forensic Personhood:** Persons are inherently forensic entities; they are the kinds of entities which can be held morally responsible for their actions, for which present and future self-regarding prudential concern is rational, and more generally for which a range of practical and ethical judgments are appropriate.[1]

The forensic conception has some standing as a (perhaps *the*) central conception of personhood, evidenced by how contested attributions of personhood (to chimps, dolphins) so often revolve around attributions of moral status. Its most

[1] See Shoemaker (2007 : 317–8) for a fuller accounting of the prudential and ethical concerns.

famous expression is found in Locke, who held that 'person' is a forensic term, 'appropriating actions and their merit; and so belongs only to intelligent agents capable of a law, and happiness, and misery' (1694: 50–1). Letting 'f-person' pick out persons on the forensic conception, person essentialism should thus be understood as *f-person essentialism*. This makes person essentialism a substantive thesis.

The standard way to reject f-person essentialism is *animalism* (e.g. Snowdon 1991; Olson 1997, 2007). The animalist claims that we are most fundamentally and essentially (human) animals with biological persistence conditions.[2] However, she does not deny that we are f-persons. We are indeed. It's just that we are only *contingently* f-persons. We are f-persons in the way we are professors, firemen, nurses, or New Yorkers. We can fail to be firemen and New Yorkers, and we can likewise fail to be f-persons. And the persistence conditions of f-persons *qua f-persons* are not biological, but psychological. In this respect, and perhaps in this respect only, the animalist agrees with neo-Lockean psychological continuity theorists (e.g. Lewis 1976; Parfit 1984; Shoemaker 1984): while the neo-Lockean is wrong to endorse f-person essentialism, she is right to say that f-persons persist *qua f-persons* via psychological continuity.[3] Not all f-persons are, or have to be, animals. Some could be angels or even wholly digital entities—beings who are not fundamentally biological. But *we* are animals, and essentially so. And even if my practical and ethical concerns as an f-person tracks my biological continuity as an animal in the normal course of things, the rationality of such concerns won't hold *in virtue of* any biological continuity. As Olson (1997: 70) tells us, 'the relations of practical concern that typically go along with our numerical identity through time are closely connected with psychological continuity...the Biological Approach has an interesting ethical consequence, namely that those practical relations are not necessarily connected with numerical identity.'[4]

But animalism is not the only way to reject f-person essentialism. Another way—and the focus of the present essay—is:

Subject Essentialism

(i) We are fundamentally and essentially subjects of experience, who persist via relations of phenomenal continuity.

[2] Some animalists have argued that we are animals, yet have psychological (or some other non-biological) persistence conditions (see Bailey 2015 for discussion). Here I am concerned with what might be thought of as 'classic animalism'. On this view, you do not go with your cerebrum. I'll lump the view that we are most fundamentally organisms under the (classic) animalist umbrella.

[3] Compare: animalists reject 'student essentialism', but can agree that the persistence conditions of students *qua students* involves being in school. Thanks to Matt Duncan for this example.

[4] See also DeGrazia (2002). For psychological continuity theories that include f-person essentialism, see, e.g. Nozick (1981) and Baker (2000).

(ii) We are f-persons, but only contingently so, and whether we are is dependent on the kind of mentality we achieve beyond being phenomenally conscious.

I proceed on the assumption that (i) is sufficient for animalism to be false. In this way, whatever the subject of experience is, it is not fundamentally an animal (Duncan 2015). This implies some measure of neutrality on the nature of subjects, but this is no embarrassment. After all, animalism is in the same boat; animalism *itself* does not commit us to any non-trivial claims about the metaphysical nature of animals. Just consider the gulf between Olson's (2007: 27–9) variant of animalism and Thorton's (2019) hylomorphic variant. That (i) is sufficient for the falsity of animalism comports with our assumption that animalism is more than the claim that we *are* animals—something more than what Johnston (2016: 121) calls the 'pure predicative view'. Animalism says that we are *fundamentally* animals.[5]

But I won't argue for (i). I will assume it. Instead, my focus will be on clause (ii). This constitutes the rejection of f-person essentialism, and a positive claim about what it takes to be an f-person. Specifically, I'll contend that f-personhood and f-personal persistence require a suite of non-phenomenal mental capacities akin to those posited by classic neo-Lockean psychological continuity theorists, along with the non-phenomenal form of consciousness that underlies them, namely, our ability to *access* our experiences. This ability is the manifestation of Block's (1995) *access consciousness*, a functional notion of consciousness. A mental state is access-conscious just when it is well-placed for rational control of reasoning, action, and report (in creatures so capable), and available to cognitive processing more generally. That a mental state is accessible does not imply that every aspect of that state is actually accessed, but it is by being actually accessed that your state (and its content) takes root within the broader cognitive economy of your mind. This distinction—being accessible to cognitive systems (and so access-conscious) versus being accessed by those systems—will prove crucial in what follows.

The approach taken up here is unique in at least two ways. The first is dialectical. Consider Olson and Witt's (2020: 2) starting point, what they call *Lockean person essentialism*. On this view, persons are beings with 'special mental powers…something that is intelligent and self-conscious in a certain way, where self-consciousness is the ability to think about oneself in the first person'. This idea can be traced to Locke's claim that a person 'is a thinking intelligent Being, that has reason and reflection, and can consider it self as itself, the same thinking thing in different times and places…' (1689/1694: 2.27.9). Since this

[5] Put another way, ANIMAL is our most basic substance concept. See Olson (1997: 36). It is doubtful that ANIMAL (or ORGANISM) is a substance concept (Nichols 2010). This is especially concerning for the animalist since other arguments for animalism, like the 'Too Many Thinkers Argument', are only arguments for the pure predicative view. See Duncan (2021) for discussion.

notion of personhood emphasises mentality, let's call such persons *m-persons* and the corresponding essentialist thesis *m-person essentialism*. Prescinding from Olson and Witt's exact mental characterization, the proponent of Subject Essentialism will contend that all f-persons are m-persons. But this is technically a further claim that requires argumentation. I supply it. That is, the tack here takes certain practical and normative features as basic, from which it will be *argued* that for us to have these features—for us to be f-persons—we must be (something like) m-persons, i.e., beings with 'special mental powers' that outstrip the purely phenomenal. So while both f-person and m-person essentialism count as 'Lockean' insofar as their popularization is traceable to Locke, they differ in their Lockean starting points.[6] It is one thing to say that f-persons are mental in nature, it is another thing to specify which aspects of mentality are required for f-personhood.

Second, Subject Essentialism occupies a unique place within the theoretical landscape. With animalism, Subject Essentialism rejects f-person essentialism. But against animalism, it denies that we are fundamentally animals. With psychological continuity theories, Subject Essentialism concurs that we are fundamentally mental entities. But against psychological continuity theories, it denies that we are fundamentally f-persons, and that we persist via (non-phenomenal) psychological continuity. With phenomenal continuity theorists (Dainton & Bayne 2005; Strawson 2009; Dainton 2008; Duncan 2015, 2019), Subject Essentialism concurs that we are fundamentally subjects of experience who persist via phenomenal continuity.

With respect to where Subject Essentialism breaks with traditional phenomenal continuity theories, it is fairly clear that at least some traditional theorists are interested in f-personhood, even if this point is not made as explicit as one might like.[7] For instance, Dainton & Bayne (2005: 556–7) speak of having 'reasons to be afraid' where only phenomenal continuity is present. This suggests that they are concerned with f-persons. In describing a radical form of virtual reality in which, due to a mishap, everything save for your stream of consciousness is destroyed, Dainton (2008: 20) speaks of surviving and being able to enjoy a 'eventful and rewarding life', thus creating an option one ought *want* in lieu of being unplugged and dying.[8]

[6] I use 'person' and 'person essentialism' where neutrality is needed.

[7] Duncan (2015, fn. 19) is an exception; he's explicit that he does not use 'person' for f-person.

[8] McMahan (2002)—who holds something like an embodied variant of the phenomenal continuity view—has a take on egoistic concern that is closest to what follows, especially in its emphasis on psychological unity and time-relative interests. However, McMahan (*ibid.*: 67–8, 79) also tells us that the physical and functional continuity of the brain regions that realize consciousness is necessary and sufficient for a 'minimal' degree of egotistic concern. The sufficiency claim is at odds with Subject Essentialism. Consciousness might be essential to being a *welfare subject*, i.e., being the kind of entity that can be better or worse off. Yet my being a welfare subject is not enough to ground the rationality of my having future-directed *concern* for my welfare level.

The deeper point I want to advance, though, is this. The traditional phenomenal continuity theorist tells us what we are fundamentally (subjects of experience) and how we persist (via relations of phenomenal continuity). If they intend to be telling us what it is to be an f-person, then their theory is false. If they are not, then I contend that they are nonetheless missing something important. They are missing *what matters* in survival. A relation matters in survival just when it is in virtue of that relation that one will (e.g.) be justified in having prudential egocentric concern at all—concern for how one's future survival turns out. Parfit (1984) argued that identity is not the relation that matters, at least as such. It matters only insofar as it confers psychological continuity, permitting cases where what matters is preserved in the absence of literal numerical survival. I argue that the traditional phenomenal continuity theorist is giving us the converse scenario. While phenomenal continuity might be necessary and sufficient for our numerical identity over time, phenomenal continuity is not the relation that matters either, at least not all on its own. Experiences are great, but they are not enough.

I'll begin by articulating the core motivations for the phenomenal continuity view. Given our running assumptions, this discussion will be brief, and I won't have anything unique to add regarding the nitty gritty of just how this continuity might work.[9] However, care will be taken to set out the view in a way that does not presuppose f-person essentialism. This will provide us with a sharper depiction of where Subject Essentialism and traditional phenomenal continuity views overlap, and a proper setting from which the argument for clause (ii) can be set out. That argument will include, amongst other things, a closer look at the forensic nature of personhood, and an account, borrowed from Schechtman (2014), that allows us to demarcate the conditions under which we can fail to be f-persons.

II. STREAM DIVERTERS AND BRAIN-STATE SCANNERS

The *phenomenal approach* to personal persistence is a close cousin of the psychological approach to personal persistence. Whereas the latter appeals to *psychological continuity*, constituted by non-occurrent 'psychological' states like beliefs, desires, intentions, character traits, hopes, fears, and memories bound by causal relations, the former appeals to *phenomenal continuity*, constituted by temporally ordered and unified experiences and phenomenal states.

In defence of the phenomenal continuity theory, Dainton & Bayne (2005) has us consider two devices: a stream diverter and a brain-state transfer device. A stream diverter redirects the flow of consciousness from one brain to another, preserving phenomenal continuity while having no effect on psychological

[9] This has been covered extensively elsewhere. See Dainton (2008) and Duncan (2019).

continuity. Transferring the being's psychology alongside her consciousness requires a brain-state scanner. With this distinction in mind, Dainton and Bayne invite us to consider the following scenarios:

> **Scenario One** (S1): Brain state transfer devices and streamal diverters are used in tandem: Smith's psychology and stream of consciousness are transferred to your brain, your psychology and stream of consciousness are transferred to Smith's brain (2005: 556).
> **Scenario Two** (S2): Brain state transfer devices are used to transfer Smith's psychology to your brain, and your psychology to Smith's brain. Streamal diverters are not used; phenomenal continuity is unaffected; your stream of consciousness continues to be sustained by your original brain, and Smith's by his (2005: 557).

S1 is a full mental transfer; S2 is a partial mental transfer. For Dainton and Bayne, the lesson in each case concerns person preservation: S1 shows that 'the combination of phenomenal continuity and psychological continuity is person preserving, and that physical continuity by itself is not', while S2 shows that 'the combination of phenomenal continuity and physical continuity is person preserving, and that psychological continuity by itself is not' (2005: 558).

The implication here is that phenomenal continuity is person preserving but psychological continuity is not. We want to distinguish between being person preserving in the neutral, trivial sense (where 'person' refers to just whatever we are essentially) and being f-person preserving, so that phenomenal continuity theories do not build in the latter. Here are three theses to that end:

T1: The combination of phenomenal continuity and psychological continuity is f-person preserving.
T2: F-personal persistence obtains in virtue of phenomenal continuity.
T3: F-personal persistence obtains in virtue of psychological continuity.

On Subject Essentialism, we are fundamentally subjects of experience, and only contingently f-persons, where f-personhood is a matter of our having a suite of non-phenomenal mental capacities (including memory formation), and the cognitive access to experiences that underlies them. Subject Essentialism entails T1, ¬T2, and T3. This leaves open whether there are f-persons that are *not* beings fundamentally like us—that is, whether there are f-persons who are *not* subjects of experience that persist via phenomenal continuity. Either way, the rejection of T2 for T3 makes the reference to phenomenal continuity in T1 incidental; for beings fundamentally like us, there is no psychological continuity without phenomenal continuity (*x* at *t* cannot be the same f-person as *y* at *t** without *x* and *y* being the same subject of experience), even if there might be phenomenal continuity without psychological continuity. Subject Essentialism is a theory about what *we* are, and it says that we are beings who can—given the right development—be f-persons. Whether things *other than us* can also be f-persons is left aside.

Since T2 and T3 will come up repeatedly, I'll give them more recognisable names. Let T2 be F-Person Preservation$_{PHEN}$ and T3 be F-Person Preservation$_{PSY}$. I'll now argue that F-Person Preservation$_{PHEN}$ is false.

III. DISTINGUISHING SUBJECTS AND F-PERSONS

III.1. The argument

The connexion between our ordinary concept of personhood and practical and normative concerns runs deep. But as Schechtman (2014: 14–6) points out, we can think of the forensic nature of personhood in two different ways. One way is when we consider a person as what Schechtman calls a *forensic unit*: 'a suitable target about which particular forensic questions can be raised and judgments made' (*ibid.*: 14). Another way is to consider a person as a *moral self*, or to 'see the limits of a single person as set by the very actions and experiences for which she is in fact held rightly accountable' (*ibid.*: 15). Schechtman makes the distinction particularly vivid with the following case:

> [W]e can...break judgments of accountability into two steps. First there is a basic question of attribution–when Mother comes into the living room to find Dick, Jane, and Spot next to the broken lamp there is an initial question of which one of them actually came into contact with the lamp and knocked it over. Suppose that Mother determines that the proximate cause of the lamp's being knocked over was contact with Jane's body. Another step is still needed, most people feel, before we judge that Jane should be held fully culpable. Perhaps Jane willfully pushed the lamp over during a tantrum, or was careless again despite several warnings, but perhaps she had one of her seizures or fainting spells and flailed into the lamp, or perhaps Spot ran in and knocked her over, or perhaps she is only eighteen months old and has no idea of the consequences of putting her hands on the lamp. (*ibid.*: 14).

When Mother chooses Jane as the child to question, Mother is considering Jane qua forensic unit. Here the boundary conditions of a single person are defined by those actions (e.g. knocking over the lamp) for which she is *potentially* responsible, and those events and experiences (e.g. being punished for knocking over the lamp) towards which it is *potentially* rational to have egoistic concern (*ibid.*: 19). It is a further matter whether Jane is *rightly* held responsible for knocking over the lamp, and whether having egoistic concern about being punished for doing so *is* rational. For Schechtman, such actions, events, and experiences define the boundaries of the moral self. If x and y are the same moral self, then x and y are the same forensic unit. The converse entailment, however, can fail to hold.

Taking Schechtman's distinction on board, the friend of Subject Essentialism cannot rest easy with simply showing that phenomenal relations are not sufficient to ground *appropriate* practical and ethical judgments associated with

a moral self. She must also show that they are not sufficient to even ground *potentially* appropriate such judgments, associated with a forensic unit.

The argument to be run, then, looks like this:

P1: *x* and *y* are the same f-person only if *x* and *y* stand in the right forensic relations—only if *x* and *y* are the same forensic unit.
P2: It is possible for *x* and *y* to be phenomenally continuous without standing in the right forensic relations, i.e., without *x* and *y* being the same forensic unit.
∴ F-Person Preservation$_{PHEN}$ is false: the relation *being phenomenally continuous with* is not identical to, nor grounds, the *being the same f-person as* relation.

I'll now set out two cases in support of P2. The first will address future-directed egocentric concern. The second will address moral praise and blame.

III.2. In defence of P2

III.2.1. Case one: egocentric concern

Let's begin with:

> **The Afterlife:** John dies. A being wakes up in the afterlife. That being is phenomenally continuous (in the right way) with John. So, by the phenomenal continuity theory, that being *is* John. In this version of the afterlife, phenomenal continuity and psychological continuity automatically diverge. John is *not* psychologically continuous with his pre-afterlife self. In fact, John cannot form any new non-phenomenal psychological states: he has no beliefs, desires, intentions, or future memories. Beyond consciousness, John is an empty mental shell, unable to cognitively access his ongoing stream of experience, and unable to report on what he sees, hears, smells and tastes.

The Afterlife does not presuppose much about phenomenal consciousness. It is consistent with a wide range of theories. Having run through this point first will prove beneficial when we set out how *The Afterlife* bears on P2.

Recall our earlier distinction between phenomenal consciousness and access consciousness. Being able to access our experiences is what makes them available for cognitive processing—being able to think, remember, and report on them (Block 2011). There are many questions one might ask about access consciousness, but these need not detain us here. The present point is this: although most parties would grant a conceptual distinction between phenomenal consciousness and access consciousness, whether they can dissociate is a point of dispute.

For example, Tye's (2000: 63) PANIC theory—'PANIC' standing for poised, abstract, non-conceptual intentional content—claims that a mental state is phenomenally conscious only if it is suitably poised to impact downstream cognitive systems. Tye does not frame things in terms of access consciousness, but that's precisely what *being poised* is; so on Tye's theory, there are no

phenomenally conscious states that are not also access conscious. However, *The Afterlife* does not rule out PANIC theory. This is because *being poised*, and access consciousness, are dispositional, functional role properties (Tye 2003: 276), and we can happily grant that in this odd version of heaven, John's phenomenally conscious mental states are poised. They simply don't *actually* result in any higher-level cognitive responses. In *The Afterlife*, there is no uptake by higher-level consuming systems, which is essential for its inhabitants (like John) to create the mental states across time that seem to underwrite f-personhood. So John's conscious states are *accessible*, but they are never *actually accessed*, perhaps because God masks the relevant disposition. It's like a Ming vase whose fragility is never manifested.

A similar storey applies to Prinz's (2012) AIR theory. According to Prinz, phenomenally conscious states are attended intermediate level representations (hence 'AIR'). Central to Prinz's theory is *working memory*—memory that subserves ongoing action, categorization, and behaviour in the broadest sense. Working memory is a canonical case of cognitive access, and is the 'leading edge' of long-term memory (Block 2011). Now for any normal subject, if you hear (e.g.) a trumpet, your perceptual system will register the sound, and to report on what you heard, the perceptual representation produced by that system must not only be accessible to working memory, but also encoded in and thus actually accessed by working memory. In *The Afterlife*, John can hear a trumpet—he can have an auditory experience of a trumpet—but he cannot report on what he hears. Why is that? On one extreme, we can envision views where the perceptual representation doesn't even need to reach accessibility to be conscious; the representation is conscious but nonetheless inaccessible to working memory. Such a view is consistent with *The Afterlife*, since the further step (whatever it is) between perceptually registering the sound and being accessible to encoding in working memory might simply be severed.[10] But *The Afterlife* is also consistent with less extreme views like Prinz's, which just requires that conscious states be *accessible* to working memory (2012: 97). And that's enough to explain John's inability to tell you what he heard, since doing that requires actual access and encoding.

Are there any theories of phenomenal consciousness that require our mental states be accessed, as opposed to merely being accessible? It's tempting to read *higher-order theories* of consciousness this way. It is tempting because they claim that we are *aware of* our conscious mental states, in addition to being *aware with* those mental states. Being aware of our mental states is indeed a way of accessing those states. But the higher-order theoretic family is a big tent, and it comes with no definitional commitment that this form of access constitutes encoding in working memory, or the kind of access relevant to f-personhood.

[10] Block's (2007) discussion of visuo-spatial extinction subject 'GK' suggests something like this 'inaccessibility' view, although it's unclear whether this is Block's considered position.

For example, on Rosenthal's (2005) higher-order thought ('HOT') theory, the core commitment is just to an consciousness-constituting assertoric HOT.[11] On Hellie's (2007) acquaintance theory, the core commitment is to a primitive higher-order acquaintance relation. So the point made about the PANIC and AIR theories can be made here too.[12]

Back now to P2. We are considering whether *The Afterlife* eradicates forensic connexions. In particular, we are asking: while numerically identical to John in virtue of phenomenal continuity, does the being in *The Afterlife* remain a proper object of John's future-directed prudential concern? In light of P1, a 'no' answer speaks to the falsity of F-Person Preservation$_{PHEN}$.

It might be initially tempting to answer in the affirmative. John would care about what happens to the being in *The Afterlife*, the thought goes, because (by stipulation) the being would be *John*. And there are facts about the exact character of his trajectory: if there is wine in *The Afterlife*, John will be able to drink it and taste its flavour, and if there is a nice beach, he will be able to lie on it and feel his feet in the sand. And no doubt, these are in some sense good things. Nonetheless, this initial reaction warrants scepticism. To see why, it will help if we tease out what things might be like for John post his Earthly-death. The picture is not rosy.

John is a cognitive husk. He does not want anything. He has experiences, but he does not want more of ones he might (before his Earthly-death) have thought of as pleasant, or less of ones he might (before his Earthly-death) have thought of as unpleasant. He does not believe that his wife on Earth loved him, or that he is undergoing experiences now. He can't form beliefs about his perceptual environment. He can see a sunset, but because he cannot access his experience of the sunset, he cannot form beliefs about the perceptible qualities—the lush pinks and yellows—represented by his experience that he might otherwise have relished. He remembers nothing of Earth, and now lives entirely within the specious present.

It is harder to envision what things might be like for John from a first-person perspective for a simple reason: we *can* do all these things and we are *not* cognitive husks. Still, a decent approximation is feasible. Here's a case from Block (1995 : 234) that, with some adjustments, should help. You are engaged in an all-consuming conversation with a friend. During this conversation, an extremely loud pneumatic drill is running outside. Around noon—the conversation having been conducted for about fifteen minutes already—you finally notice that someone is running the drill. Now it's implausible to deny that you literally did not hear the drill prior to noticing it. You suffer from no

[11] Indeed, there are versions of HOT theory where the question of 'actual access' doesn't get off the ground, since the HOT itself is best understood as the conscious state. See Gottlieb (2020, 2021).

[12] A potential exception is the Global Workspace Theory ('GWT'). It is not obvious that GWT (Dehaene & Naccache 2001) was intended as a theory of phenomenal consciousness. The good news is that *qua theory of phenomenal consciousness*, GWT is false (Block 2007, 2011).

auditory deficits, and the drill was loud enough. So since you heard the drill, you experienced it. But, prior to noon, even if your experience of the drill was accessible, it was never actually accessed, perhaps because your attention was not appropriately allocated due to the intense conversation. I suggest that John's life is the *limit* of Block's pneumatic drill case. That is, in the envisioned scenario, *all* of his experiential states are akin to the state of hearing the drill pre-noon. So to get a sense of what (e.g.) *The Afterlife* might be like from the first-person perspective, we simply take what it is like to hear the drill pre-noon, and let all of one's experiential states be like *that*.

Further insight can be had by comparing something that would normally be thoroughly enjoyable—a luxury cruise, say—with trying to do that very same thing in *The Afterlife*. So let's imagine, then, that upon arriving in *The Afterlife*, John wakes up on the cruise. He is ushered throughout the ship, having visual experiences of the beautiful open water—visual experiences that, if he had his full suite of cognitive mechanisms, he would recognise as pleasant, want more of, and fondly remember. John is then taken to a Michelin 3-star on-board restaurant, where he feasts upon lobster tails, porterhouse steak, and bananas foster—resulting in gustatory experiences that, if he had his full suite of cognitive mechanisms, he would recognise as pleasant, want more of, and fondly remember. But he cannot do *any* of these things. He can have a photograph taken of himself at dinner, but he would have no idea who is in the picture, or whether the person in the picture had the bananas foster. Indeed, if John was asked whether he was tasting bananas foster, he wouldn't be able to say either way. This is because the requisite access would be missing; the representation that constitutes gustatory experience, though accessible John's reporting mechanisms, would not be encoded in them.

Taken together, these glimpses into *The Afterlife* suggest that, from the perspective of John prior to his Earthly death, the being in *The Afterlife* is not a proper object of his future-directed egocentric prudential concern. That object might *be* John, but that is not enough. Nor is it enough that John will have experiences. Although these experiences would in one sense be *John's*—they occur 'in' him—there is another sense in which they would not be *for John* in virtue of their functional isolation. Masking the disposition underlying access consciousness *masks the experiences from John*. And this holds not just for any single experience, but sets of experiences over time, as the masking will preclude John's ability to 'mentally stitch' together his experiences as happening to the same 'I'—viz. *him*. So if experiences in *The Afterlife* will not be *for John*, it's hard to see why John would have any preferences whatsoever about what happens after his Earthly death.

One could object: despite the functional isolation of John's experiences, each individual experience would be intact, with their normal phenomenal character. For instance, couldn't John still have intense pain in *The Afterlife*? Wouldn't it be rational for pre-afterlife him to prefer that he doesn't have such experiences?

I concede that the immediate impulse for most who consider this matter would be to say that *of course!* such future-directed egocentric concern is rational. We reflect on what pains are typically like for normal subjects like us, and how they bear on the rest of our mental lives, and based on these factors, we then transfer these reflections over when considering how things must be for someone like John, concluding that his life will be *subjectively horrible*. This comparison, however, is far too easy. The impulse should be resisted.

Let's start by distinguishing between pains and painfulness. Pain is a kind of sensation. In the paradigmatic case, they are sensations associated with some threat to the physical integrity of the body. The phenomenal character of pain includes facts about felt location, shape, and duration. Pains also have a distinctively negative affective component. They are painful; they hurt. Cluster headaches, for instance, feel terrible. And it is not just bodily pains that are painful; some emotions are painful, and we can be so hungry that it hurts (Klein 2014 : 53–4). Not all pains are like this, though. As Klein (*ibid.*) notes, the sensation associated with regular postural adjustment doesn't hurt, but it is still a pain, possessive of other familiar pain-markers like motivational force. Pain asymbolics might present another, more extreme case: subjects whose pains never hurt (Bain 2014). The upshot here is just that, what pre-afterlife John might rationally want to avoid is not pains *per se*, but painfulness—be it bodily pains that hurt, or painful experiences more generally.

With this distinction in mind, we might say that it would be rational for S at t_1 to not want her future self S* to have a painful experience at t_2 only if S has reason to believe that S* will not want to have a painful experience at t_2. There are two things to note about this necessary condition. First: it is not trivially satisfied even for ordinary subjects, given masochistic pleasure. For masochists, pains hurt. They just *like* them. These pains are *pursued*, and not just due to contextual factors (Klein 2014 : 44). The expression 'hurts so good' is not oxymoronic.

Second: it strikes at the very grounds of future-directed egocentric concern. Such concern is often based on assumptions about what our future selves will want or desire, or what we think will make us happy more generally. I save money for retirement because I know that, in the future, I will want money and prefer a situation where I have some to one where I don't, all things considered.

The problem here should be fairly clear: in *The Afterlife*, masochist or not, this necessary condition is simply not met. John's cannot prefer, desire, believe, remember, or exercise any other cognitive capacity. John will not be the kind of entity capable of *wanting to not* have a life filled with painful sensations. In *The Afterlife*, John has no desires to be frustrated. He has no preferences that can fail to be met. Even if John was a masochist pre-afterlife, there would be no luxuriating in his afterlife-pains. Whether a life counts as 'subjectively horrible' plausibly requires that one has preferences and desires to begin with.

Suppose though we shift away from a desire-satisfaction model and instead assume the following sufficient condition: having a life rife with painful experiences is sufficient to make John's life subjectively horrible, and so sufficient to ground the rationality of pre-afterlife John's preference to avoid such a life. Even so, this might not cut it, since it is not at all clear that John can have painful experiences in *The Afterlife*. That John will be phenomenally conscious and phenomenally continuous with his pre-afterlife self does not entail that he will be able to have all the kinds of experience he used to have. For even if John can have (e.g.) bodily sensory pains, painfulness is a further step that might require a non-experiential source extrinsic to the pain proper. And at least two otherwise widely divergent models say it does. On desire views, pain sensations count as painful because we desire that they cease. This desire is extrinsic to the pain sensations, and the desire is not *itself* an experiential state. Yet in *The Afterlife*, John is not capable of having desires.[13] On Klein's (2015: 183–6) imperative theory of painfulness, a first-order sensation (a first-order state with body-involving imperative content) is painful just when it is the object of a higher-order imperative. This higher-order imperative is not itself an experiential state, but *confers* painfulness on the first-order sensation. Yet John cannot undergo anything but experiential states in *The Afterlife*. So he won't have experiences that are painful.[14]

What if instead *The Afterlife* effected qualitative duplication—duplication at a purely phenomenological level—to the extent possible, but we also stipulated that this duplication was based on a case where John's life was by-and-large good experientially? Would this not create a life just like an ordinary one? It might have some painful experiences, but on the whole it seems worth having.[15]

It is a mistake to think that qualitative duplication would result in anything approximating an ordinary, good life. For one thing, the point made about painfulness also applies to pleasure; without *de re* desires for a gustatory experience of bananas foster, John might not find it pleasurable (Heathwood 2007). But let's wave this and take a different tack. Consider the character of Leonard Shelby from Christopher Nolan's classic 2000 film *Memento*. Shelby is not quite like John; he retains some cognitive faculties, but suffers from anterograde amnesia, a condition in which one is unable to form new memories. Despite this, Shelby is phenomenologically whole. He has no experiential deficits in virtue of his amnesia.

[13] See Heathwood (2007). This point also holds on Brady's (2018) variant of the desire view, where painfulness attaches to a *relation* between the pain (the sensory state) and the desire; you cannot have the relation if one of the relata (the desire) is missing.

[14] Klein intends this account of painfulness to apply to *all* forms of pain, so the claim here isn't merely that (e.g.) John's stubbed toe won't be painful.

[15] Dainton's (2008: 20) remarks on his 'VR-4' scenario suggest this view.

Various things happen to Shelby that, from an experiential perspective, we would normally consider to be unpleasant. Yet there are also things that happen to him which are horrible due to the amnesia; independent of getting beaten up by some thugs, his post-condition life is not one anyone would desire. So we can abstract away from those purely experiential details and imagine a scenario in which Shelby had a life that was just like his precondition life experientially-speaking but still inflicted with anterograde amnesia. I submit that such a life is preferable to no life at all, but just barely. Yet now remember: John's condition in *The Afterlife* is vastly worse than Shelby's. For Shelby, long-term memory loss is just forward-looking; for John it's backward-looking and forward-looking. And worse still, if the contents of John's experiences are not encoded in working memory, John's experiential life, such as it is, would be wholly inert. That such a life would be desirable is hard to envision.

III.2.2. Case two: responsibility and blame

What about judgments of responsibility and blame? Consider:

> **The Murder:** John hates Bill. As a consequence, much of John's mental life is pre-occupied by feelings of rage towards Bill. Above all, John wants Bill dead, and intends to kill him. Just after he carries out his plans, a brain-state transfer device wipes John's psychology clean, while retaining phenomenal continuity. Post-procedure John is not psychologically continuous with pre-procedure John. There is no connexion, causal or otherwise, between pre-procedure John's intention to kill Bill, and his later doing so. Nor does post-procedure John remember wanting to kill Bill. In fact, John's new psychological life is antithetical to his past life: he is immovably happy and pleasant, and sees forgiveness–not retribution for whatever Bill may have done–as essential to his moral fiber. So not only does John not remember who Bill is, he is shocked at that thought that he would have killed him.

Unlike *The Afterlife*, in *The Murder*, John can form new memories, beliefs, desires, and intentions. However, these states bear no connexions to pre-procedure John. Moreover, given John's new sunny disposition, their contents will tend towards being very different. Whereas pre-procedure John would tend to erupt in anger at the perception of even the smallest slight, post-procedure John would shrug it off. Post-procedure John not only has no memory of killing Bill, he can't conceptualise being the kind of man who killed Bill. John doesn't recognise what it is to desire to perform such actions, beyond in the most abstract sense.

These considerations suggest that post-procedure John isn't even a *potential* target for blame for his pre-procedure actions. Although he is numerically identical to pre-procedure John—and so *the thing* that committed murder—they are distinct f-people. He is not simply a different f-person in a metaphorical sense, as when your child returns home from college pronouncing that she is now into 'goth'. Here, being a 'new man' is not a matter of a transitional,

psychological metamorphosis. By stipulation, post-procedure John has *zero* psychological connexions with pre-procedure John. The forensic unit he was no longer exists.

Contrast this with a case where no brain-state transfer device is used, but prior to the murder John was instead fed mind-altering drugs against his will. Here we might resist saying that post-intoxication John is *rightly* held responsible. But he is surely a *potential* target for blame. That's to be expected: since psychological continuity doesn't require unbroken psychological connectedness—recall Parfit (1984) here—pre and post-intoxication John are the same forensic unit.

But suppose we add that it's *John* who has arranged for the brain-state transfer device to wipe his prior psychology clean. Wouldn't he then be blameworthy?[16] No. For notice: this scenario is not relevantly different from a scenario in which John had not just arranged for his prior psychology to be wiped clean, but also for him to be incapable of forming any new psychological states, as in *The Afterlife*. In those cases, I take it, it would be wrong to blame John. Yet why would things be different if, as in *The Murder*, John *is* able to form new psychological states? That's *a* difference, but not a difference that makes a difference. John might understand what 'murder' means, but he will not remember murdering anyone, and will sincerely insist that he has no idea how one could *commit* murder in practice. New John is like most us for 'necrophilia' and *being* a necrophile.

All of this is of course consistent with our intuitions that a terrible wrong was committed in Bill's murder, and our natural desire for justice and that *someone* be punished. Post-procedure John is the closest reasonable candidate, but that does not mean that he *is* to blame and *should* be punished. Indeed, theories of punishment seem to suggest that he *ought not* be punished. On a preventive-leaning theory of punishment, punishment is unjustified. With John's new psychology, there is nothing to prevent. Retributive-leaning theories, while perhaps more friendly to punishing John, don't obviously deliver that verdict either since it is at best unclear whether John can be a desert subject despite having no psychological connexions to his former crime-committing self.[17] And lest we worry too that a deliberate-psychology-decanting-scenario might give anyone some bad ideas about how to escape punishment, note that the present account predicts that such a decision, while not resulting in literal death, is quite close if not worse than literal death. So few would pursue it. And in any event, we already have suicide bombers.

[16] Many thanks to an anonymous referee for suggesting this alteration to *The Murder*.

[17] Ryberg (2012) argues that retributivists can construe psychology-changing interventions as a legitimate form of punishment. So if John had arranged for his psychological clean slate (with the ability to acquire a new one), he might have punished himself, albeit in a perhaps unjustified way.

III.3. F-personhood preservation

Such is the case for P2, and thus against F-Person Preservation$_{PHEN}$. But we can now take it a step further. We have a case for F-Person Preservation$_{PSY}$:

P3: F-personhood and f-personal persistence is mental in nature.
P4: If f-personhood and f-personal persistence is mental in nature, then either F-Person Preservation$_{PHEN}$ or F-Person Preservation$_{PSY}$ is true.
P5: F-Person Preservation$_{PHEN}$ is false.
∴ F-Person Preservation$_{PSY}$ is true.

It might be objected that P4 presupposes a bad taxonomy, consisting of a clean break between the phenomenal and the psychological. This issue is at least partially terminological. The sort of states relevant to F-Person Preservation$_{PSY}$ are *non-phenomenal*, and indeed must be, since they need to be *standing* states. Your memory of your tenth birthday party does not go away simply because you are not currently thinking about it, or fall asleep.[18] Thus we can recast the consequent of P4 in terms of phenomenal versus non-phenomenal (but still mental) forms of f-personal persistence. Of course, these non-phenomenal psychological capacities are entangled in various ways with the experiential aspects of our mental lives. It might be that having a memory of your tenth birthday party involves being able to token an episodic experiential memory, not just a propositional judgment that my birthday party had cake and lots of friends. This experiential manifestation, however, would not be *constitutive* of F-Person Preservation$_{PSY}$.

While this disjunctive syllogism is not nothing, we should not overlook the direct line we have to the conclusion. Reflecting on *The Afterlife* and *The Murder*, it is clear that the kinds of mental states relevant for f-personhood appealed to by standard neo-Lockean psychological theories–i.e., beliefs desires, intentions, and memories, braided together in overlapping causal chains–and the forms of cognitive access that underlie them, are precisely what's missing. In other words, *The Afterlife* and *The Murder* provide direct support for F-Person Preservation$_{PSY}$. In *The Murder*, John presumably *would* be morally blameworthy for killing Bill if he had psychological states that were connected to his past self. So he no longer is the same f-person. In *The Afterlife*, John's experiences are *accessible* to the relevant consuming systems that *would* produce the relevant psychological states. Yet without *actually being in* these states, no *actual* mental scaffolding is created upon which an f-personal identity can be hung. So here John is not an f-person at all. Memory, the ability to plan and think about your future—things essential to f-personhood—requires access, not just accessibility.

[18] This is why the *bridge problem*—explaining how we survive dreamless sleep—holds for phenomenal continuity theories. For solutions, see Dainton & Bayne (2005) and Gustafsson (2011).

In sum, Subject Essentialism says that we are fundamentally and essentially subjects of experience, and we persist via phenomenal continuity. As Dainton (2008) puts it, you cannot escape your own stream of consciousness (Fig. 1):

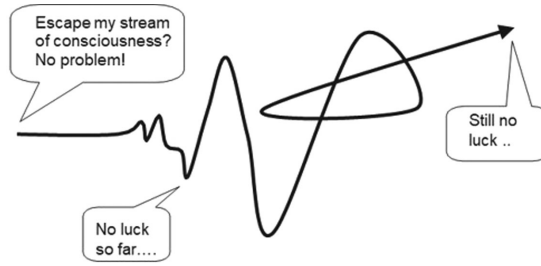

Figure 1. "Attempting the Impossible" (Dainton 2008: 26).

Yet being a subject of experience is not sufficient for being a f-person. The picture advocated here countenances (Fig. 2):

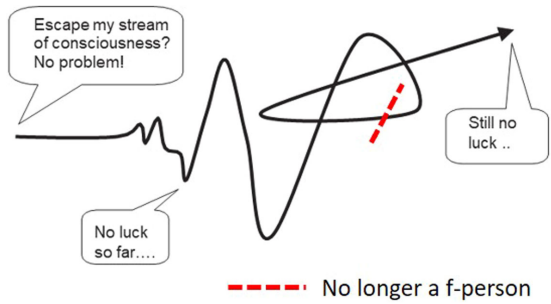

Figure 2. Attempting the Possible.

On Subject Essentialism, you can fail to be an f-person if you lose the ability to access your conscious experiences, and you can fail to be the same f-person if your beliefs, desires, intentions, and memories are not connected in the right way.

IV. CONSCIOUSNESS AND WHAT MATTERS

Subject Essentialism tells us what we can and cannot fail to be. We cannot fail to be subjects; we cannot survive zombification. Yet we can fail to be f-persons. An overarching lesson of the preceding is that this may be worse than literal death, insofar as being an f-person, not a mere subject, is what matters.

This is not idle theoretical speculation. Subject Essentialism has bite in real life. Take the vegetative state (VS). It is not implausible to think that despite

sleep–wake cycles and eye movements, if a subject is in a VS, they are wholly phenomenally unconscious. But in a functional magnetic resonance imaging study of a 23-year-old female subject, Owen *et al.* (2006) found that their patient, although lacking in any overt behavioural markers of consciousness (e.g. verbal report), showed the same neural activity as healthy controls when given tasks involving mental imagery. Owen *et al.* take this to be evidence of phenomenal consciousness in the VS:

> These results confirm that, despite fulfilling the clinical criteria for a diagnosis of vegetative state, this patient retained the ability to understand spoken commands and to respond to them through her brain activity, rather than through speech or movement. Moreover, her decision to cooperate with the authors by imagining particular tasks when asked to do so represents a clear act of intention, which confirmed beyond any doubt that she was consciously aware of herself and her surroundings. (Owen *et al.* 2006: 1402)

Assume that Owen *et al.*'s diagnosis is correct.[19] What does this mean for their patient? Subject Essentialism tells us: unless Owen *et al.* can also show that their patient *accesses* her conscious states, she has lost what matters. It is technically a further step to deny her full moral status, but it is not implausible that f-personhood bears on this too (cf. Levy & Savulescu 2009). If John had little reason to care about what happens to his future self in *The Afterlife*, we should think that, barring future recovery, we would have little reason to care about what happens to our concious-but-vegetative future selves.[20]

REFERENCES

Bailey, A. (2015) 'Animalism', *Philosophy Compass*, 10: 867–83.
Bain, D. (2014) 'Pains that Don't Hurt', *Australasian Journal of Philosophy*, 92: 305–20.
Baker, L. R. (2000) *Persons and Bodies: A Constitution View*. Cambridge, UK: CUP.
Block, N. (1995) 'On a Confusion about a Function of Consciousness', *Brain and Behavioral Sciences*, 18: 227–47.
—— (2007) 'Consciousness, Accessibility, and the Mesh between Psychology and Neuroscience', *Behavioral and Brain Sciences*, 30: 481–54
—— (2011) 'Perceptual Consciousness Overflows Cognitive Access', *Trends in Cognitive Sciences*, 15: 567–75.
Brady, M. (2018) 'Painfulness, Desire, and the Euthyphro Dilemma', *American Philosophical Quarterly*, 55: 239–50.
Dainton, B. (2008) *The Phenomenal Self*. Oxford: OUP.
——Bayne, T. (2005) 'Consciousness as a Guide to Personal Persistence', *Australasian Journal of Philosophy*, 83: 549–70.
DeGrazia, D. (2002) 'Are We Essentially Persons? Olson, Baker, and a Reply', *Philosophical Forum*, 33: 81–99.
Dehaene, S. and Naccache, L. (2001) 'Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework', *Cognition*, 79: 1–37.

[19] Unsurprisingly, this is not uncontroversial. See Shea & Bayne (2010).

[20] I wish to think Marya Schechtman, two anonymous referees for *The Philosophical Quarterly*, and especially Matt Duncan for helpful comments and discussion.

Duncan, M. (2015) 'I Think, Therefore I Persist', *Australasian Journal of Philosophy* 93: 740–56.
—— (2019) 'A New Argument for the Phenomenal Approach to Personal Persistence', *Philosophical Studies*, doi: https://doi.org/10.1007/s11098-019-01297-x.
—— (2021) 'Animalism is Either False or Uninteresting (Perhaps Both)', *American Philosophical Quarterly*, 58: 187–200.
Gottlieb, J. (2020) 'On Ambitious Higher-Order Theories of Consciousness', *Philosophical Psychology*, 33: 421–41.
—— (2021) 'The Higher-Order Map Theory of Consciousness', *Australasian Journal of Philosophy*. https://doi.org/10.1080/00048402.2020.1861038.
Gustafsson, J. (2011) 'Phenomenal Continuity and the Bridge Problem', *Philosophia*, 39: 289–96.
Heathwood, C. (2007) 'The Reduction of Sensory Pleasure to Desire', *Philosophical Studies*, 133: 23–44.
Hellie, B. (2007) 'Higher-Order Intentionalism and Higher-Order Acquaintance', *Philosophical Studies*, 134: 289–324.
Johnston, M. (2016) 'Remnant Persons: Animalism's Undoing', in S. Blatti and P. Snowdon (eds.) *Animalism: New Essays on Persons, Animals, and Identity*, 89–127. Oxford: OUP.
Klein, C. (2014) 'The Penumbral Theory of Masochistic Pleasure', *Review of Philosophy and Psychology*, 5: 41–55.
—— (2015) *What the Body Commands: The Imperative Theory of Pain*. Cambridge: MIT Press.
Levy, N. and Savulescu, J. (2009) 'Moral Significance of Phenomenal Consciousness', *Progress in Brain Research*, 177: 361–70.
Lewis, D. (1976) 'Survival and Identity', in A. Rorty (ed.) *The Identities of Persons*. Berkeley, CA: University of California Press.
Locke, J. (1689/1694) *An Essay Concerning Human Understanding, (The Clarendon Edition of the Works of John Locke)*, P. Nidditch (ed.). Oxford: OUP.
McMahan, J. (2002) *The Ethics of Killing: Problems at the Margins of Life*. Oxford: OUP.
Nichols, P. (2010) 'Substance Concepts and Personal Identity', *Philosophical Studies*, 150: 255–70.
Nozick, R. (1981) *Philosophical Explanations*. Cambridge, Massachusetts: Harvard University Press.
Olson, E. (1997) *The Human Animal: Personal Identity without Psychology*. New York: OUP.
—— (2007) *What Are We? A Study in Personal Ontology*. New York: OUP.
——and Witt, K. (2020) 'Against Person Essentialism', *Mind*, 129: 715–35.
Owen, A. *et al.* (2006) 'Detecting Awareness in the Vegetative State', *Science*, 313: 1402.
Parfit, D. (1984) *Reasons and Persons*. Oxford: OUP.
Prinz, J. (2012) *The Conscious Brain*. Oxford: OUP.
Rosenthal, D. (2005) *Consciousness and Mind*. Oxford: Clarendon Press.
Ryberg, J (2012) 'Punishment, Pharmacological Treatment, and Early Release', *International Journal of Applied Philosophy*, 26: 231–44.
Schechtman, M. (2014) *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. Oxford: OUP.
Shea, N. and Bayne, T. (2010) 'The Vegetative State and the Science of Consciousness', *The British Journal for the Philosophy of Science*, 61: 459–84.
Shoemaker, S. (1984) 'Personal Identity: A Materialist Account', in S. Shoemaker and R. Swinburne (eds.) *Personal Identity*, 67–132. Oxford, UKS: Blackwell.
Shoemaker, D. (2007) 'Personal Identity and Practical Concerns', *Mind*, 116: 317–57.
Snowdon, P. (1991) 'Personal Identity and Brain Transplants', *Royal Institute of Philosophy Supplement*, 29: 109–26.
Strawson, G. (2009) *Selves*. Oxford: OUP.
Thorton, A. (2019) 'Disembodied Animals', *American Philosophical Quarterly*, 56: 203–17.
Tye, M. (2000) *Consciousness, Color, and Content*. Cambridge, MA: MIT Press.
—— (2003) 'On the Virtue of Being Poised: Reply to Seager', *Philosophical Studies*, 113: 275–80.

Texas Tech University Texas, USA. E-mail: joseph.gottlieb@gmail.com