# Table of Contents

1. **Holism, Narrative, and Paradox: New Criteria for Settling Disputes in Personal Identity**  
   Jaron J. Cheung  
   1–20

2. **Is Anyone on First? Sport, Agency, and the Divided Self**  
   Jeffrey P. Fry  
   21–34

3. **Persons, Person Stages, Adaptive Preferences, and Historical Wrongs**  
   Mark E. Greene  
   35–49

4. **Naturalizing Persons: A Reply to Lynn Rudder Baker**  
   Eric Kraemer  
   51–74

5. **Brutal Personal Identity**  
   Peihong (Karl) Xie  
   75–94
Holism, Narrative, and Paradox: New Criteria for Settling Disputes in Personal Identity

Jaron J. Cheung
The State University of New York at Buffalo

Biography
Jaron J. Cheung is currently a PhD student in philosophy at the University at Buffalo, SUNY. His research interests are in metaphysics (especially personal identity, perception, ontology, and time), social and political philosophy, ethics and moral psychology, applied ontology, philosophy of race, philosophy of education, and philosophy of religion.

Acknowledgments
I am indebted to David Hershenov for providing extensive and insightful comments on multiple drafts of this paper. Moreover, I owe my gratitude to those at the 2023 Persons Conference, hosted by the Center for Cognition and Neuroethics at the University of Michigan—Flint, where this paper was first presented, and Peihong (Karl) Xie and Finn Wilson for providing helpful feedback during the final stages of revision.

Publication Details

Citation
Holism, Narrative, and Paradox: New Criteria for Settling Disputes in Personal Identity

Jaron J. Cheung

Abstract
This paper introduces three new criteria that a theory of personal identity ought to satisfy: (1) material holism, (2) narrative unity, and (3) narrative integrity. Material holism guards against the undesirable consequence of positing the person as part and existentially distinct from the organismal whole, of which it is dependent and interconnected. Narrative unity ensures that continuity between the beginning, middle, and end of a human life is sufficiently accounted for. Narrative integrity secures fidelity and congruence between each part and the whole, the whole to each part. Jeff McMahan’s Embodied Mind Account (EM) fails to satisfy each of these. On McMahan’s account, human persons and human organisms are distinct entities, human persons come into existence after its human organism, and human persons may go out of existence before their human organism. Moreover, fetuses, infants, the congenitally severely cognitively impaired, those with severe dementia, and the comatose are non-persons. A theory of personal identity that incorporates holism and narrative can provide a better explanation of human existence, life and death, and the identity paradox of dicephalic twins. If accepted, EM must either be rejected or ameliorated, and the new criteria ought to be incorporated in contemporary research of personal identity.

Keywords
Personal Identity, Criteria, Material Holism, Narrative Unity, Narrative Integrity, Animalism, Embodied Mind Account, Dicephalic Twins

What does it mean to be human? Debates on what kind of beings humans are essentially or fundamentally have primarily terminated between two rival traditions in contemporary philosophy: animalism and psychologism. Broadly speaking, animalism represents a cluster of views that identify human persons with human animals—or that we are essentially human organisms—and psychologism, a cluster of views that identify human persons with a psychological criterion—or that we are essentially psychological beings. In Jeff McMahan’s, The Ethics of Killing: Problems at the Margins of Life, McMahan issues the challenge of dicephalus, a case of twins conjoined below the neck and sharing what seems to be one body, to conclude that human persons are distinct from their human organism. Upon analysis, McMahan determines that animalism cannot sufficiently account for the identity paradox, a puzzle of the relationship between the person and the animal found in the case of dicephalic twins, Abigail and Brittany Hensel, and recommends his Embodied Mind Account to settle the metaphysical problems of what human persons are and their persistence conditions over time. As a consequent,
McMahan’s account leads him to conclude that: (1) human persons are distinct from human organisms, (2) human persons come into existence after its human organism, (3) human persons may go out of existence before their human organism, and (4) early abortion is permissible up to 20 weeks because there is no human person that is harmed.

In this paper, I contend that McMahan’s Embodied Mind Account fails to adequately capture what humans are and erroneously reduces persons to mere psychological capacities, i.e. the minimal capacity for thought and sentience. On McMahan’s account, fetuses, infants, the congenitally severely cognitively impaired, those with severe dementia, and the comatose are non-persons. Although McMahan’s account embraces brain continuity (i.e. physical and minimal functional continuity of cerebral structures) as a criterion for personal identity over time, it simultaneously prescribes narrative discontinuity (i.e. discontinuity between the beginning, middle, and end of human existence). As such, McMahan’s account provides a fragmented and incomplete picture of human life.

In order to show this, I introduce three new criteria that a theory of personal identity ought to satisfy: (1) material holism, (2) narrative unity, and (3) narrative integrity. Material holism guards against the undesirable consequence of positing the person as part and as existentially distinct from the organismal whole, of which it is intimately dependent and interconnected. Narrative unity ensures that continuity between the beginning, middle, and end of a human life is sufficiently accounted for. Narrative integrity secures fidelity and congruence between each part and the whole, the whole to each part. (I will detail these criteria in a later section of this paper.) McMahan’s Embodied Mind Account does not satisfy the aforementioned and thereby leads him to conclude that there exists multiple overlapping entities, i.e. human persons overlapping human organisms, and moral prescriptions for abortion grounded by psychologism. A theory of personal identity that incorporates holism and narrative identity can provide a better explanation of human existence, life and death, and the paradox of the Hensel twins. As a result of what follows, if what I show in this paper is true, McMahan’s Embodied Mind Account and subsequent moral prescription for early abortion must either be rejected or ameliorated.

It is important to note that McMahan also appeals to the cerebrum transplant thought experiment to argue for the intuition that persons are not identical to their organism. As it is beyond the scope of this paper, I will not treat McMahan’s cerebrum transplant thought experiment here.

If this paper is successful in resolving the challenge of dicephalus in a way that suggests persons are animals or something else, McMahan’s cerebrum transplant scenario
would still have need to be sufficiently addressed. I will maintain this thesis in five parts: (I) The Challenge of Dicephalus: Abigail and Brittany Hensel, (II) The Embodied Mind Account, Existence, and Abortion, (III) Inverse Excurses—The Challenge of Craniopagus: Krista and Tatiana Hogan, (IV) Holism, Narrative, and Personal Identity, and (V) Closing, Paradox, and Hensel Twins Revisited.

I. The Challenge of Dicephalus: Abigail and Brittany Hensel

There is another challenge to the view that we are organisms that need not appeal to examples drawn from science fiction but instead focuses on an actual, though extremely rare, condition known as dicephalus. Dicephalus (from Greek roots, meaning “two-headedness”) occurs when a human zygote divides incompletely, resulting in twins conjoined below the neck. In dicephalic twinning, as in other forms of twinning, it is clear that there are two people. In a case featured in a recent issue of Life magazine, Abigail and Brittany Hensel present a spectacle of two heads sprouting from a single torso; yet no one doubts that they are separate and distinct little girls. Each has her own private mental life and her own character, each feels sensations only on her own side of the body, and each has exclusive control over the limbs on her side… But, although Abigail and Brittany are two different persons, there seems to be only one organism between them. If so, then neither girl is identical with that organism. For they cannot both be identical with the organism, as that would imply that they were identical with each other, which they are not. (McMahan 2002, 35)

McMahan issues the challenge of dicephalus to those who countenance the animalist view that human persons and human animals are identical. Departing from brain and cerebrum transplantation cases, McMahan invokes the real-life case of Abigail and Brittany Hensel to make the claim that human persons are distinct from their human organism. Although the Hensel twins “have two hearts and two stomachs, they share three lungs, have a single liver, a single small intestine, a single large intestine, a single urinary system, and a single reproductive system” (McMahan 2002, 36). These “organs are packaged together within a single rib cage and function together in a harmoniously coordinated manner” (McMahan 2002, 36). Thus, the Hensel twins having two heads
arising out of a single body, according to McMahan, is an example of two distinct human persons in one human organism.

For those who hold that we are essentially human organisms, in order to determine what might be the most plausible explanation of the Hensel twins personal identity status, McMahan considers what he thinks are the only three possible options: (1) dicephalic twins constitute a single organism and therefore can be at most one person—a person with a divided mind, (2) dicephalic twins constitute a single organism with two distinct minds, and (3) dicephalic twins are actually two distinct though overlapping organisms (2002, 35–36). McMahan asserts that (1) and (2) are unacceptable for the same reasons: that both claims deny that either Hensel twin can be a separate and independently existing thing (2002, 36). The third claim, which McMahan believes is most promising, is not satisfactory as it is like “the claim that a plane with duplicate control mechanisms for a pilot and copilot is really two distinct but overlapping planes” (2002, 37). McMahan thinks that in cases of dicephalus, in opposition to the view that there are two distinct overlapping organisms, there is a single biological life that supports the existence and thus the lives of two distinct persons (2002, 37). As such, McMahan believes that the challenge of dicephalus as presented by the Hensel twins, seems to be a “clear case in which there are two persons who coexist with and are supported by a single organism” and “that there are two persons present, one per cerebrum” (2002, 39). It is here that McMahan makes the further conclusion that as the dicephalic twins are not a different kind of entity from ourselves, or that a different account of personal identity applies to them, we are not essentially organisms either (2002, 39). We too are parts of organisms. That is, we non-twins stand to organisms in the same relationship as the dicephalic girls. Thus, McMahan rejects the view that we are essentially human organisms and moves into considering the Psychological Account of Personal Identity—the view that we are essentially psychological beings.

II. The Embodied Mind Account of Egoistic Concern, Existence, and Abortion

The Embodied Mind Account of Egoistic Concern (hereafter Embodied Mind Account) was developed out of careful analysis of and an amelioration of the Psychological Account of Personal Identity (McMahan 2002, 39–88). According to McMahan’s account, we are essentially embodied minds (2002, 68). The criterion of personal identity across time on this account is physical and minimal functional continuity of the parts of
brain that produce thought, where “physical continuity of an organ such as the brain requires either the continued existence of the same constituent matter or the gradual, incremental replacement of the constituent matter over time” and “functional continuity involves the retention of the brain’s basic psychological capacities” (McMahan 2002, 68). What is meant here by basic psychological capacities is the capacity for consciousness and the different capacities that come with consciousness, e.g. pain, pleasure, etc. For what provides “the basis for egoistic concern about the future, is continuity or sameness of consciousness”, that is, “the continuity of the capacity for consciousness, so that the renewed appearance of conscious states following a period of unconsciousness is always the reappearance of the same consciousness, or the same mind” (McMahan 2002, 67). Thus, “the relation that is constitutive of identity—sufficient physical and functional continuity of the areas of the brain in which consciousness is realized in order for those areas to retain the capacity to support consciousness—is both a necessary and a sufficient condition of a minimal degree of rational egoistic concern” (McMahan 2002, 79).

Rational egoistic concern is important for the Embodied Mind Account because it is a requisite for McMahan’s Time-Relative Interests Account, with which he uses in part to determine the goodness and badness of life and death and the permissibility of abortion. For McMahan, rational egoistic concern about some event within one’s own future life is a function of two factors: “first, the value, positive or negative, that the event would have for one at the time when it would occur, and second, the extent to which the prudential unity relations would hold between oneself now and oneself at the later time when the event would occur” (2002, 79–80). Prudential unity relations are characterized in part by psychological connectedness and continuity, which McMahan identifies with organizational or structural continuity, i.e. the “preservation of those configurations of tissue that underlie the connections and continuities among the contents of an individual’s mental life over time” (2002, 68; 74). Organizational and structural continuity, prudential unity relations, and psychological connectedness and continuity are not required criteria for McMahan’s account of personal identity, but they are important for grounding rational egoistic concern and identifying what matters in a human person’s life, including one’s time-relative interests. One’s interests, in the sense that McMahan uses it, is to “have an interest in something for one’s well-being to be engaged with it” (2002, 80). The present time-relative interests of an individual “are what one has egoistic reason to care about now” and “are always, as the label is intended to suggest, relativized to one’s state at a time” (McMahan 2002, 80). In order to determine the strength of one’s present time-relative interests in the possibilities of one’s future life, we would undertake a discounting operation where the value of future events that one would have within
one’s life at the time they would occur are “multiplied by a number (either I or a positive fraction) representing the strength of the prudential unity relations between oneself now and oneself at those times when the events would occur” (McMahan 2002, 80). Having established a basic foundation and terms for personal identity and in identifying what matters in a human person’s life according to the Embodied Mind Account, let us now turn to what a person is and when they come into existence.

To be a person, “one must have the capacity for self-consciousness and perhaps, a mental life with a high degree of unity” (McMahan 2002, 90). “Person” is a term that refers to what we essentially are in a generic way (McMahan 2002, 90). McMahan is unclear about when a person generally arises along the timeline of a human organism’s development. However, McMahan takes the human person to be a phase sortal, a kind to which an individual may belong through only part of its history (2002, 7; 24). Prior to a person’s existence, along the timeline of a person’s development, there is what might be called the mindless, i.e. the fetus at 0-20 weeks, and the minimally minded, i.e. the fetus 20 weeks through birth to infancy. The person arises at some time after the minimally minded is developed to a sufficient degree which would satisfy the conditions of personhood. After a person phases out of existence, it is also possible for there to be the minimally minded and mindless post-person. For example, the minimally minded post-person could be a result of progressively worsening Alzheimer’s or brain damage, while the mindless post-person could be a result of severe dementia or those who have become irreversibly comatose.

What is important to note is that the fetus at 0-20 weeks and the congenitally severely cognitively impaired never acquired the status of persons and those with severe dementia and the comatose, such as those in a persistent vegetative state or those who suffered brain trauma, have lost their status as persons because they either never acquired or no longer have the capacity for consciousness or lack a mental life with a high degree of unity. Counter to common intuitions about what qualifies as a person, the Embodied Mind Account says that the aforementioned are not persons and possess an inferior moral status. McMahan alludes to this in his preface:

Among those beings whose nature arguably entails a moral status inferior to our own are animals, human embryos and fetuses, newborn infants, anencephalic infants, congenitally severely retarded human beings, human beings who have suffered severe brain damage or dementia, and human beings who have become irreversibly comatose.
These are all beings that are in one way or another “at the margins.” (2002, vii)

There are at least two ways that McMahan identifies which beings possess a moral status inferior to our own and their subsequent moral treatment: (1) the Embodied Mind Account’s criterion of personal identity, i.e. the presence of physical and minimal functional continuity of cerebral structures, to determine moral status and (2) the Time-Relative Interests Account to determine moral treatment.

On the first strategy McMahan employs, if a being doesn’t satisfy the criterion of personal identity, then there is no person to kill. McMahan writes that “we do not begin to exist until our organisms develop the capacity to generate consciousness” (2002, 267). Thus, those beings that do not possess consciousness, in particular, fetuses at 0-20 weeks, are not persons and possess an inferior moral status. (McMahan notes that consciousness, at earliest, may develop at 20 weeks or roughly 5 months and that early abortion is thus performed prior to 20 weeks [2002, 268].) McMahan elucidates: “An early abortion does not kill anyone; it merely prevents someone from coming into existence. In this respect, it is relevantly like contraception and wholly unlike the killing of a person. For there is, again, no one there to be killed” (2002, 267).

As such, early abortion is permissible because there is no one there to be harmed by killing. On the Embodied Mind Account, this logic extends to other human beings that do not qualify as persons, such as the congenitally severely cognitively impaired, the severely demented, and the comatose, which may seriously offend common intuition and sensibility.

But what if there are human beings that do possess minimal consciousness and qualify at least as minimally minded? Although, we will not delve deeply into McMahan’s second strategy which utilizes the Time-Relative Interests Account, it would be helpful for us to see the logical conclusions of the Embodied Mind Account. McMahan observes that “there are some human beings whose psychological capacities are no more advanced than those of certain animals”: (1) fetuses at 20 weeks and on and infants, (2) those with acquired cognitive deficits (e.g. those who have suffered brain damage or dementia), and (3) congenitally severely cognitively impaired human beings (2002, 204). Due to their “rudimentary cognitive and emotional capacities, human beings of all three types have a comparatively weak time-relative interest in continuing to live” (McMahan 2002, 204). The pregnant woman with a fetus at 20 weeks and on may have a later abortion because the fetus’s time-relative interest is so minimally tied to their future, that the mother’s time-relative interests in not being pregnant outweighs the fetus’s (McMahan
Members of the second and third groups of human beings—those with acquired cognitive deficits and the congenitally severely cognitively impaired—have such weak relations with themselves in the future that the “Time-Relative Interest Account implies that it would be no more seriously wrong, other things being equal, to kill a human being of one of these two types than it would be to kill an animal with comparable psychological capacities… Very few people will find this a welcome conclusion” (McMahan 2002, 205).

As mentioned from the outset, we will delimit our inquiry to the case of early abortion which permits an abortion of a fetus at 0-20 weeks. Recall that according to McMahan, “these abortions merely prevent someone like you or me from existing… there is no one there to be killed” (2002, 268). As 99 percent of all abortions are performed prior to 20 weeks (McMahan 2002, 268), focusing our analysis on early abortion and McMahan’s theory of personal identity which grounds its permissibility will be our task. What is of interest to us is whether McMahan’s Embodied Mind Account of Identity sufficiently captures what human persons are. For if McMahan’s criterion for personal identity is wrong, then it follows that the Embodied Mind Account must either be rejected or ameliorated and consequently, its prescription for the permissibility of early abortion must be as well. Moreover, I suspect that this would have implications for McMahan’s Time-Relative Interest Account and the aforementioned conclusions regarding the inferior moral status of human beings on the margins and the permissibility of their being killed (although I will not treat this in this paper). Even further, a pressing concern is the personal identity and moral status of fetus and infants, the congenitally severely cognitively impaired, those with severe dementia, and the comatose. If the Embodied Mind Account is correct, then it follows that because consciousness and a high degree of mental unity is not present in these beings, then these beings are not persons. This claim, offensive to many, goes against ordinary intuitions about the personal identity and moral status of such beings. As such, it is a welcome task to critically analyze the challenge of dicephalus and the Embodied Mind Account. To begin, we must briefly consider an inverse challenge to the Embodied Mind Account, the challenge of cephalothoracopagous janiceps.
III. Inverse Excurses—The Challenge of Craniopagus: Krista and Tatiana Hogan

The Hogan girls, Krista and Tatiana... share part of their brains and this leads to what seems to be a sharing of some thoughts. If one is pricked by a needle drawing blood, the other winces. If one drinks something delicious, the other verbally expresses her pleasure... The girls’ relatives have even suggested that their shared thoughts go beyond the sensual. If one is looking at the television while the other’s line of vision doesn’t include the television, the latter might still laugh at something that stimulated only the eyes of the former. It doesn’t seem that the girls ever suffer ambiguous self-reference, each twin unaware whether she is Tatiana or Krista. There are instead two minds engaged in a sort of “telepathic eavesdropping.” One would say “ouch” when the other was pinched out of sight because the message would be sent via the shared parts of their brains. (Hershenov 2013, 204–205)

The challenge of craniopagus is an inverse case of the dicephalic conjoined twins, Abigail and Brittany Hensel. Where dicephalic twins are conjoined below the neck and share an organism, craniopagus twins are conjoined above the neck at the cranium, with some cases sharing part of their brain. For our purposes, we are interested in the latter, craniopagus conjoined twins that share part of their brains. As McMahan invoked the real-life case of dicephalic conjoined twins, Abigail and Brittany Hensel, let us briefly consider the real-life case of craniopagus twins, Krista and Tatiana Hogan.

The Problem of Too Many Thinkers is often charged against psychological views of identity by animalists. For example, animalist Eric Olson “maintains that if the person is spatially coincident but numerically distinct from the animal, then provided that the person can use its brain to think, so too can the physically indistinguishable animal” (Hershenov 2013, 203). Thus, according to the Problem of Too Many Thinkers the psychological identity theorist seems to posit multiple thinkers in the same organism. McMahan thinks differently, however, and replies that the Embodied Mind Account avoids the problem because “it is the brain-sized person who truly thinks, while the animal thinks only in a derivative sense in virtue of having a thinking proper part” (Hershenov 2013, 203). In the case of dicephalic conjoined twins, McMahan’s Embodied Mind Account may prove advantageous because it would identify dicephalic twins as two brain-sized persons in one organism, potentially solving the personal identity paradox.
Yet, if the Hogan twins successfully present the groundwork for an inverse challenge to the psychological identity theorist, then it follows that there is warrant for the inverse claim that human persons are not brain-size parts of human organisms.

The Hogan twins share a thalamus, which connects to both of their brainstems. Although the thalamus works directly in tandem with the activity of the cerebrum and is believed to be involved with the activity of consciousness, the Hogan twins do not directly share a partially overlapping cerebrum. In order to present a plausible case of craniopagus twins that would Pose a Problem of Too Many Thinkers for psychological identity theorists like McMahan, David Hershenov tweaks the Hogan twins case to a different case of conjoined twins with partially overlapping cerebra (2013, 204–205). In Hershenov’s adjusted case example of “Hogan-like” twins, the conjoined twins qualify as sharing partially overlapping cerebra which renders them spatially coincident and being reduced to a condition of sharing all their thinking parts (2013, 204–205). In the adjusted Hogan-like twins case, the unshared parts are destroyed and each thinker becomes smaller and spatially coincident with the other. It is here the Embodied Mind Account of Identity encounters an inverse challenge and problem: How many thinking persons are there even though they share the same neurology and generate consciousness from the same shared cerebra? (Hershenov 2013, 204–205). Recall that the Embodied Mind Account’s criterion for personal identity is physical and minimal functional continuity of cerebral structures. In the case of Hogan-like twins, the Embodied Mind Account would have to admit that there are two thinking persons that have their mental life and consciousness generated by the same shared cerebral structures (Hershenov 2013, 205). Yet, if there is only one shared cerebral structure, how can two distinct thinking persons emerge from the same neurology? It appears that not only does a Problem of Too Many Thinkers arise, the adjusted Hogan-like craniopagus twins case also provides McMahan with an inverse-like problem that he challenges the animalist with in the case of dicephalic conjoined twins.

If the Embodied Mind Account’s criterion for personal identity admits that there are two thinking persons that share the same cerebral structures, then a similar inverse charge of the kind that McMahan issues against animalists in the challenge of dicephalus also arises against the Embodied Mind Account theorist. Recall the challenge of dicephalus reformulated as a conditional: If there are two distinct persons (two distinct cerebrums, one per person) and one shared human organism, then persons are existentially distinct from the human organism. Against the Embodied Mind Account theorist, an inverse-like challenge of craniopagus (i.e. the adjusted Hogan-like twins case) formulated as a conditional is: If there are two distinct bodies, two distinct minds exemplified
by diverging brain activity (one per body), and one shared cerebral structure (where unshared parts are destroyed and each thinker becomes spatially coincident with the other), then there are two thinkers present that arise from the same cerebral structure. The Embodied Mind Account theorist would have to admit that in the case of the Hogan-like twins, there are two thinking persons that arise from the same shared cerebra, which is an inverse problem that McMahan charges against the animalist in the case of dicephalic twins. Recall that McMahan was not convinced that the dicephalic twins could be two distinct overlapping organisms and instead thought the most plausible view was that there are two distinct persons that coexist in one organism. A similar problem exists for the Embodied Mind Account theorist—either Hogan-like twins exist as two distinct persons that arise out of overlapping cerebral structures or there are two distinct persons that coexist in one shared cerebrum. Yet, as we have already identified that this poses the Problem of Too Many Thinkers (and it would be true for each case), it seems that in being charitable to the Embodied Mind Account theorist, the most plausible alternative would be that Hogan-like twins are a case of a single shared cerebrum with a divided mind. This seems implausible, however, because with this explanation the individuated minds of both Hogan-like twins would be lost, thereby losing the force behind the theory that we are essentially embodied minds that arise from individuated physical and minimal functional continuity of cerebral structures. This shows that at the very least, there are plausible reasons that warrant suspicion of the Embodied Mind Account’s criterion of personal identity, as well as reason to seek alternative accounts of personal identity that better preserve our intuitions about what we are. Moreover, although this brief excurses does not solve the identity paradox of dicephalic and craniopagus twins (and it does not claim to), the adjusted case of Hogan-like twins weakens the advantage McMahan claims over animalism in his appeal to intuitions about personal identity and the dicephalus. To appease the dissatisfaction that has left us wanting, we must consider new criteria for settling disputes in personal identity that can better point us in the right direction.

IV. Holism, Narrative, and Personal Identity

In this section, I will briefly set forth a preliminary account, although not comprehensive, of narrative identity and its parts relevant to our task. Against this backdrop, we will be able to grasp what both holism and narrative have to offer in developing three new criteria for settling disputes in personal identity: (1) material holism, (2) narrative unity, and (3) narrative integrity. The following subsections will
have focused evaluative questions that any theory of personal identity must satisfy to adequately capture what and who we are.

Narrative identity in relation to questions of personal identity and characterization possess four features that are of interest to us: (1) humans are story-telling animals (MacIntyre 2007, 216), (2) the lives of persons are narrative in form (Schechtman 1996, 93–135) (3) narrative identity is co-constructed individually and communally (Schechtman 2014, 89–109), and (4) narrative identity may render paradox intelligible within a cohesive, continuous, and unified whole (Ricoeur 1992, 113–168). Story-telling is a praxis central to human existence. So fundamental to human praxis is the telling of narrative that it is arguable that perhaps all of theory, including philosophical inquiry on personal identity, is mediated through it.

Narratives have a beginning, middle, and end and a human’s narrative identity is co-constructed between the individual (i.e. self-creating reflexive consciousness and utterance) and the individual’s community (i.e. third-person identifying referential utterance) (Ricoeur 1992, 50–55). Marya Schechtman identifies this co-constructive practice between the individual and the individual’s community by identifying three features of narrative construction: (1) self-narratives are generated from the first-person perspective, (2) an identity-constituting narrative is not just a story you have about yourself but also the stories others tell about you, and (3) those without the wherewithal to narrate their own lives (e.g. infants and those with cognitive deficits) can be identified through narratives created by others (2014, 103–104). In practice, this looks like a mother and father speaking to a fetus in the womb expressing excitement for their eventual arrival, addressing infants and young children as if they will eventually possess forensic capacities though they do not have them yet, and treating dementia patients and those that are comatose as the continuation of a particular narrative (e.g. visiting dementia and comatose patients, overseeing their care, supplying them with their favorite things from the past) (Schechtman 2014, 104–105). On the narrative view, the boundaries of what constitutes personhood may be extended in a much more egalitarian sense than what the Embodied Mind Account allows for.

Finally, narrative identity has the unique capability of rendering contradiction and paradox intelligible within a cohesive, continuous, and unified narrative whole. According to Paul Ricoeur, what marks and is characteristic of all narrative composition is discordant concordance (1992, 141). Narrative succeeds in bringing together the discordant properties of one’s life into a unified concordant whole. What are contradictory facts in one’s story may be rendered intelligible when considering the cohesive, continuous, and unified narrative whole. This does not mean, however, that all narratives are true. A
narrative might be partly or wholly fiction. Nonetheless, the mechanism of narrative is robust enough to make intelligible paradox in a human person’s life.

With this brief introduction to narrative identity, let us now turn to considering the three new criteria that any theory of personal identity ought to satisfy.

IV.I Material Holism

*Does this theory provide a sufficient account of the dependent and interconnected parts of the whole, such that the parts cannot exist independently of the whole?*

Let us define *holism* as the theory that dependent and interconnected parts of a whole cannot exist independently of the whole. There are at least two kinds of holism that would be good for our purposes to identify as possible criterion: *narrative holism* and *material holism*. Narrative holism is concerned with the parts of a story that are dependent and interconnected to the whole story. Material holism is concerned with the parts of a material being that are dependent on and interconnected to the whole being. Material in this sense are all the biophysical matter that constitutes a being. We will be concerned with the latter, *material holism*. Paraphrased then with material holism in mind, our evaluative question becomes: *Does this theory provide a sufficient account of the dependent and interconnected material parts of the whole material being, such that the material parts cannot exist independently of the whole material being?*

McMahan attempts to explain the relationship between the person and organism as mere part to the whole. In an analysis of two case analogies, (1) a tree that grows a particular limb and (2) blowing a horn in a car, McMahan concludes that “a whole (the organism) has certain properties by virtue of having a part (the mind or person) that has those properties” (McMahan 2002, 92). McMahan writes:

> Suppose, for the sake of comparison, that over a certain period of time the only part of a tree that grows is a particular limb. When this limb grows, the tree grows. The tree grows by virtue of having a part that grows. A property of the part—growth—is in this instance necessarily a property of the whole. There are thus two things that are growing: the limb and the tree of which it is a part. Similarly, when I blow the horn in my car, the horn makes a noise but so does the car. There
Cheung

are two things that have the property of emitting a noise: the horn and the car of which it is a part… These analogies help elucidate the sense in which there are two conscious entities present where I am. My organism is conscious only in a derivative sense, only by virtue of having a conscious part. (McMahan 2002, 92)

McMahan’s construal of these analogies fails to recognize that the part (the mind or person) cannot come to exist without the whole (the organism). Similarly, the limb cannot come to exist without the tree, nor can the horn (if it is electric) make a noise without being plugged into the electrical source that exists in the car. In trying to make sense of the part to the whole, McMahan does not address how the part (the mind or person) that is both dependent and interconnected to the whole (the organism) can come to exist without the organism. If the mind or person truly was in its own distinct existential category, it seems that it would be able to arise without the organism. Yet, this is not so. The mind or person cannot come to existence without the organism, nor can it be sustained without the organism.

A reformulation of the material holism criterion question for the Embodied Mind Account theorist could be: Is it possible for the person as part of the organism to arise outside the organism? As we have seen, the Embodied Mind Account theorist countenances the person as existentially distinct from the organism. However, they would also have to admit that it is not possible for a person to come into existence without the biological processes made possible by and mediated through the organism (e.g. the cellular, metabolic, cardiovascular, respiratory, and immune systems, amongst others). Such biological processes make possible the conditions for living and eventually consciousness and thinking. On the Embodied Mind Account, the material part that the person arises from, i.e. the cerebrum, is reliant on the material whole, i.e. the organism, and its processes to be developed. The cerebrum cannot be abstracted as independent from the organism, it is intimately interconnected with the whole body. Thus, the Embodied Mind Account cannot satisfy the criterion of material holism as it posits that the brain-sized person that arises from the cerebrum is independent and existentially distinct from the organism. If the Embodied Mind Account does not satisfy the criterion of material holism, then it proffers an erroneous relationship between the material part to the material whole.
IV.II Narrative Unity

Does the theory sufficiently preserve narrative continuity, cohesion, and unity between the beginning, middle, and end of human existence?

McMahan’s account allows for persons to exist after their organism comes to exist, to go out of existence in the middle of their narrative and return (e.g. those that temporarily lose basic physical and minimal functional continuity of the parts of the brain that produce thought from causes such as brain trauma or disease and regain them), and to go out of existence before their organism ceases to exist (e.g. severe dementia patients, brain trauma, the comatose). Fetuses and infants do not count as persons or one of us because they lack the consciousness and/or high degree of mental unity that would grant them personhood. On the Embodied Mind Account, it seems that we must say that if we are essentially embodied minds, then our beginnings occur much later than our organisms come to exist, we may pop in and out of existence even though our organism is still living, and we may “die” before our organism does. Yet, what if we are not essentially embodied minds, rather we are something else?

Maureen Condic writes that human organismal life begins immediately upon sperm-egg fusion. The zygote, a one-cell human organism, which forms directly after sperm-egg fusion, functions immediately to direct its own development. The zygote behaves as “an organism that is undergoing a self-directed process of maturation” (Condic 2013, 48). Condic writes: “An organism is distinct from a cell because all parts of an organism act together in a coordinated manner to preserve the life, health, and continued development of the organism as a whole” (2013, 48). In other words, the zygote is not merely a single-celled entity, nor even an eventual clump of cells. Rather, the zygote exhibits coordinated and regulatory “organismal behavior from the moment of sperm-egg fusion onward” (Condic 2013, 48). If we are essentially biological organisms, as the animalist claims, there would be greater narrative continuity, cohesion, and unity regarding what we essentially are throughout the timeline of a human life. The animalist need not worry about two entities overlapping one organism, there is simply one living and thinking animal, and would have no problems satisfying the narrative unity criterion.

The Embodied Mind Account theorist, however, has trouble with satisfying the narrative unity criterion. As the Embodied Mind Account fails to adequately satisfy the material holism criterion (by abstracting the part as independent from the whole), it then fails to adequately capture what we essentially are by claiming a false relation. If the Embodied Mind Account fails to adequately capture what we essentially are,
then the account implicitly prescribes narrative discontinuity. As we have seen in the Embodied Mind Account, narrative discontinuity is evidenced by its theoretical commitments leading it to the claim that beings such as infants, congenitally severely cognitively impaired, those with severe dementia, and the comatose lose their status as not qualifying essentially as “one of us.” This goes against common intuition, practice, and human sociality and describes a fragmented picture of human life. Yet, this is a consequence of the Embodied Mind Account and its commitments. Thus, so far we have seen that the Embodied Mind Account does not satisfy the material holism and the narrative unity criteria. Let us now consider the final criterion, narrative integrity.

**IV. III Narrative Integrity**

*Is the theory descriptively honest and the relevant parts congruent with the whole and the whole to its relevant parts?*

At first glance, this criterion might seem similar to the material holism criterion, however, this criterion is concerned with the overall integrity of the theory, that is, of whether the relevant parts are descriptively honest and congruent with the whole and the whole to its relevant parts. In the *challenge of dicephalus*, McMahan collapsed the Hensel twins individuated organs (e.g. two hearts and two stomachs, one per twin) into the narrative description that the Hensel twins coexisted and shared one harmonious organism. By making this interpretive jump to collapsing the individuated organs that belonged to each Hensel twin to a single shared organism, McMahan’s account of the dicephalic twins showcases a lack of narrative integrity. McMahan’s argument for rejecting animalism depended on the claim that the *challenge of dicephalus* really represents two distinct persons coexisting in one organism. Yet, if there are distinct organs that are not shared between the twins, then it does not follow that there really exists solely one organism. Rather, it is more appropriate to say that the Hensel twins each have their own heart and stomach, while sharing a set of organs. In maintaining that the dicephalic twins are really overlapping organisms, animalist Matthew Liao remarks: “each twin has her own stomach and heart; they have distinct brainstems and distinct spines that are only joined at the hips; and they have partially distinct organs that are united. This suggests that in fact, there are two organisms here although they are not fully independent organisms” (2006, 340).
Although Liao presents plausible reasons for dicephalic twins being overlapping organisms, it remains an open question about what the status of dicephalic twins actually are with respect to its mind, personhood, and organism. However, what could be said is that the Embodied Mind Account relies on the interpretation that dicephalic twins are a case of two persons that share a single organism and therefore narrates biological features of the dicephalic twins (i.e. distinct and unshared organs belonging to each twin) as collapsing into a singular shared organismal entity. The upshot of this strategy is that it helps McMahan’s claim that persons are distinct from their organism. The downside is that there may be features left out that are important for us to continue discourse about what the personal identity status of dicephalic twins really are.

The challenge of dicephalus is McMahan’s central real-life case example that he invokes to ground the justification for his Embodied Mind Account. Nevertheless, even without the charge of a lack of narrative integrity with regard to dicephalic twins, McMahan’s Embodied Mind Account still does not satisfy the criterion of narrative integrity. As we have seen, the criteria of material holism and narrative unity are not satisfied and therefore, as the Embodied Mind Account proffers a false relation between the part (the mind or person) and the whole (the organism), as well as implicitly prescribes narrative discontinuity, based on the final criterion, the end result is that the Embodied Mind Account does not satisfy the criterion of narrative integrity. Its parts do not align with the whole.

V. Closing, Paradox, and Hensel Twins Revisited

I hope I have shown that with a basic introduction to the inclusion of the three new criteria for settling disputes in personal identity, we may get some traction on some intractable issues. Informed by holism and narrative identity, (1) material holism, (2) narrative unity, and (3) narrative integrity as criteria can be helpful additions to help determine whether or not a theory of personal identity should be adopted. These criteria serve as standards aimed to ensure holistic alignment and narrative continuity, cohesion, unity, and integrity of the theories of personal identity in question. As we have shown beginning with the inverse challenge of craniopagus, the Embodied Mind Account theorist does not escape their own kind of challenge that they issued to animalists. In the adjusted Hogan-like craniopagus twins case, the Embodied Mind Account indeed suffers from a Problem of Too Many Thinkers. Moreover, in order to settle on its claim that persons are distinct entities from their organism, the Embodied Mind Account
must countenance a mereology that is independent from its whole, which is inherently a false relation. Brain-sized persons cannot arise on their own without the organism. Furthermore, an independent mereology that rejects the human animal as necessary to the existence of the person, leads to the consequent of an implicitly prescribed narrative discontinuity. That we are not essentially embodied minds gives us reason to consider other alternative personal identity theories that can better explain the beginning, middle, and end of a human life without positing late and fuzzy existences, as well as premature deaths. By virtue of not satisfying the first two criteria, it follows that the Embodied Mind Theorist also does not satisfy the criterion of narrative integrity. As such, the Embodied Mind Account ultimately recommends a fragmented and incomplete picture of human existence and it lacks plausibility as it relates to its account of personal identity. The Embodied Mind Account of Identity and its prescription for early abortion that it grounds must therefore be rejected or ameliorated. If an argument for early abortion is to be made, it must be made another way outside of the Embodied Mind Account’s criterion for personal identity. It is interesting to note that personal identity theories such as animalism, the hylomorphic soul theory, and the Person-Life View (Schechtman 2014, 110–138) would likely fare better at satisfying the new criteria than the Embodied Mind Account and any other psychological identity account. Another paper putting rival personal identity theories to the test would potentially prove to be a fruitful endeavor.

To return to the paradox of the Hensel twins, how should we move forward? Recall that narrative identity possesses the mechanism capable of rendering paradox intelligible in a continuous, cohesive, and unified narrative. In a narrative, the discordant contradictions that riddle a life can be brought into concordance by a unified whole. Taking a second look at the Hensel twins then, we could describe the dicephalic twins as two partially overlapping organisms that possess some of their own organs and partially share some organs that are united. In this way, we retain narrative integrity by describing what reality actually is like and we are able to then conclude that we need not make the logical conclusion that persons are existentially distinct from their human organism, for the Hensel twins are not an actual clear cut case of two heads sprouting out of a single organism with only one set of shared organs. Moreover, narrative preserves the identity individuation of each Hensel twin, while also promoting a comprehensive, holistic, and unified view of what and who we are. Holism and narrative as tools enable us to articulate a richer and fuller account of human life and thereby grant us additional pathways for getting clear on what it fundamentally or essentially means to be human.
References


Is Anyone on First? Sport, Agency, and the Divided Self

Jeffrey P. Fry
Ball State University

Biography
Dr. Jeffrey P. Fry is Professor of Philosophy at Ball State University. He holds a double major Ph.D. in Philosophy and Religious Studies from Indiana University. His recent research interests are in and at the intersection of philosophy of sport, philosophy of mind, ethics, and neurophilosophy, broadly conceived.

Acknowledgments
I would like to extend my thanks to Elizabeth Agnew for comments on a draft of this paper. Versions of this paper were presented at the “Persons Conference,” Center for Cognition and Neuroethics, University of Flint-Michigan, June, 2023 and at the group meeting of the International Association for the Philosophy of Sport, Pacific Division meeting, American Philosophical Association, Vancouver, Canada, April, 2022. I am grateful to attendees at these presentations for their feedback, and especially grateful to Nathaniel Pierce, who gave a commentary on my presentation at the Vancouver meeting.

Publication Details

Citation
Is Anyone on First? Sport, Agency, and the Divided Self

Jeffrey P. Fry

Abstract
The comedy team of Abbott and Costello performed a comic routine widely known as “Who’s on First?”. The skit exploits equivocation: specifically, use of words like “Who” “What” and “Why” as interrogatives, on the one hand, and as names of baseball fielders, on the other. This leads to dizzying, and to many, hilarious confusion. What is not disputed, however, is that someone is on first. In light of challenges from philosophy, the cognitive sciences, and personal testimonies this assumption can no longer merely be accepted at face value. At the very least, the response to “Who’s on First?” has become complex. Is someone on first? Or is it the case that no one is on first? Or are there perhaps many on first? These responses call into question the unity of the self, assumptions about human agency, and putative bases for ascribing praise and blame. I explain these challenges and examine their implications for sport. I argue that there are practical implications for both coaching and playing sports.

Keywords
Self, Sport, Coach, Athlete, Modularity, Split-Brain, Dissociative-Identity-Disorder, Plurals

Introduction
Beginning in the 1930s, the popular duo of Bud Abbott and Lou Costello performed a comedy routine widely known as “Who’s on First?” (Francis 2016). The skit exploits equivocation by using words such as “Who” “What” and “Why” as interrogatives, on the one hand, and as names of baseball fielders, on the other. This leads to dizzying and, to many listeners and observers, hilarious confusion. What is not disputed, however, is that some definite one is on first, whoever that might be. Many endorse this view, and indeed this is perhaps the default view for most of us in our unreflective moments.

But in light of challenges from philosophy and the cognitive sciences, including psychology and neuroscience, as well as personal testimonies, this assumption can no longer merely be accepted at face value. The standard, intuitive view that matches one body with one self has been variously called into question. At the very least, the answer to the question “Who’s on first?” has become more complex.
The challenges come from different directions. On the one hand, there is a spectrum of views that variously challenge the unity of the self, in some cases threatening the dissolution of the self, and in other cases the proliferation of selves. These challenges are relatively “weak” or “strong” in terms of contesting commonly held views. Among other issues, we face claims about “the modularity of mind” (Fodor 1983), “the new unconscious” (Hassin, Uleman, and Bargh 2007; Bargh 2017), the fragmented self (Levy 2018), the situated self (Ross and Nisbett 2010), and assertions by individuals that their actions don’t always represent their true selves (Eagleman 2011, 101-104). The notion of dissociative identity disorder suggests that two or more distinct personalities may be associated with one body. Neuroscientists speak about competition within the brain (Eagleman 2011, 101–150). Split-brain studies pose questions about the number of consciousnesses supported by the brain.¹ And “plurals” tell us that many persons exist as a society that shares one body (Schechter 2020).

On the other hand, we find the view that the self or the “I” is in some sense an illusion—a useful “user-illusion” (Dennett² 2017, 335–370) perhaps, a “center of narrative gravity” (Dennett 2013, 333–340), or a “strange loop” (Hofstadter 2007). It is a powerful illusion that we cannot shed, but nevertheless an illusion. Who’s on first? The counterintuitive answer is that no one is on first—at least not in the sense depicted in the “manifest image,” “the world as it seems to us in everyday life,” as opposed to the “scientific image” (Dennett 2013, 69).³

These challenges variously call into question the existence or unity of the self, personal autonomy, other assumptions about human agency, and putative bases for ascribing praise and blame. Who or what is the real or authentic self? Where does the buck stop in terms of accountability? The task of sorting out this complicated array of issues and perspectives presents daunting challenges. The theoretical landscape suggests that the old debates about personal identity and personal responsibility may need reframing.

So, who’s on first? And why should we care? Depending on the theoretical perspective that is adopted, the answer to the question, “Who’s on first?” may be someone, many, or no one. I am I. “I” am we. “I” am not. Both the correct answer and our

---

¹ See Schechter’s (2018) recent work on the topic.

² With respect to methodological considerations, see Daniel C. Dennett (2003; 2013, 341–346).

³ The distinction between the “scientific image” and the “manifest image” comes from Sellars (1962). See also Dennett (2013, 69–72).
assumed answer have practical implications for our lives in general, and specifically, for the world of sport. I provide no definitive answer to our question. However, in keeping with the sporting context of the “Who’s on First?” comedy routine, I argue in this paper that each option has implications for issues of blame, praise, and meritorious action in sport, as well as for aspiration to athletic greatness. Each option also adds complexity to the challenges of being a good coach, as well as a coachable athlete. And each option has potential ramifications for how fans might view and appreciate sport.

So, is anyone on first, and if so, who? Contrary to what the program hawker who greets you at the entrance to the ballpark would have you believe, your scorecard may not easily settle the answer to that question. Let us consider various options, some of which may overlap in certain respects.

II Someone is On First

Let us first consider the view that someone is on first, with an emphasis on one. This common view has a lengthy history. It is the view perhaps most associated with folk psychology in Western societies (though perhaps not universally), but its provenance is difficult to date. In early modern philosophy its most famous adherent was perhaps Descartes, who bequeathed his view, with all of its complications, to the subsequent history of philosophy.

In the Meditations (Descartes, [1641] 1993), Descartes announces that he is a “thinking thing” (e.g. 19, 51). That he thinks is indubitable, since, even when being deceived he must exist (Descartes [1641] 1993, 18). Being a thinking thing defines his essence. He has a body, but he is essentially a mind—an immaterial, indivisible mind (Descartes, [1641] 1993, 51, 56; Searle [2004], 8-11). This does not prevent the body and mind from interacting—with a special role given to the pineal gland (Descartes, [1649] 2021, 21–22)—though the question of how the mind and the body interact, given Descartes’ assumptions about each of them, has plagued philosophy ever since.

This thinking thing is his conscious self. So longs as, and only so long as, it exists, Descartes exists (Searle 2004, 18).4 There is no room for unconscious mental states, which might undermine the unity of the self or otherwise complicate the picture. It is consciousness that occupies center stage (Searle 2004, 21).

4. The standard joke is that Descartes went into a bar for a drink. Afterwards the bartender asked him whether he would like another drink. Descartes replied, “I think not,” and poof, he disappeared.
Aspects of Descartes’ view reverberate yet today in the popular imagination and manifest image, as well as in some philosophical accounts of the self, though with modifications. I am I. My experiences are filtered through a “dative of manifestation” (Sokolowski 2000, 65). My conscious self is the “driver’s seat,” though in more sophisticated version of this view, it may have to compete with unconscious impulses. The conscious self undergoes shifts in moods. It discovers that it has different sides. Sometimes I do not feel quite myself. I may be “off my game.” At times the spirit is willing, but the flesh is weak, and I may experience weakness of the will. I may feel pulled in different directions. But except in rare, pathological cases, there is an identity within a manifold (See Sokolowski 2000, e.g., 27–33). A sameness in difference persists. I am I, and as such, I am a responsible agent. While the self does not exemplify the attribute of divine simplicity, it displays a waxing and waning unity.

This view is economical. And when we apply it to the world of sport, it makes some things easier to understand and implement, some things more difficult, and others puzzling, if not incomprehensible.

In terms of coach-athlete interactions, it simplifies matters in some respects. A single self negotiates with another single self. Each self is perhaps complex, and sometimes obtuse, recalcitrant, or opaque, but an “I-Thou encounter,” or a “fusion of horizons,” is in principle possible. The coach must figure out what motivates the particular athlete, and the athlete must interpret and to some extent, assimilate the coach’s viewpoint. There is mentorship, teaching and learning, and skill development as coach and athlete work toward common goals. There is also accountability on the part of both athlete and coach. The athlete is a responsible agent, as is the coach. Fans heap what is, from their perspective, merited praise and blame on the individual athlete and the coach.

The idea of coaching a team adds complexity to the picture, but not exponentially so. The task is to blend these individuals into an efficient and cohesive unit that works as a unity and engages in “team reasoning” (Papineau 2017, 131–144). With large squads and numerous assistant coaches, this becomes a more complex task. And yet there is often a seemingly significant degree of cohesiveness and coordinated effort.

As noted, this view, while not simple, is, in certain respects, simple relative to other possibilities. But it is opaque in other respects. How do I explain playing in the “zone,” or how do I accomplish many athletic achievements while I am not conscious of how I am

---

5. It is said (see Lazenby 2014, 309) that Tex Winters, former assistant coach of the Chicago Bulls of the National Basketball Association, once told Michael Jordan following a game that “There’s no I in team.” Michael Jordan supposedly responded, “Yeah, but there is in win.”
executing the skills? Why do the athlete’s effort and desire wax and wane? Why does the athlete’s commitment to training vary, and the lure of temptations and diversions differ according to context or social situation? Why is an athlete sometimes a team player while at other times selfish? Why does a coach blow up, only to experience regret? Why are athletes sometimes incomprehensible even to themselves when they exhibit lapses of judgment, succumb to choking, or otherwise have subpar performances? There are perhaps responses that are consistent with this approach. But the defenses may assume forms of self-transparency and unified agency that are difficult to reconcile with the scientific image. So, let us consider a different tack.

II No One is On First

The first answer to our question “Who’s on first?” is that I am on first. To the contrary, the second response is that no one is on first—at least there is no self as traditionally understood. As we will see, this view is slippery, and as such difficult to hold within one’s mental grasp. It consists of a corpus of views that stand in a family relationship of overlapping stances. Some are more radically deconstructionist than others. In each case, an illusion is exposed. One can point here to the Buddhist doctrine of “no self.” Among Western philosophers, David Hume stated that when he cast a gaze inward, he did not discover a self, but rather only fleeting perceptions (Hume 1968, 239). Among other relevant thinkers under this large umbrella, we find Daniel Dennett, who speaks of the self both as a “user-illusion” (Dennett 2017, 335–370) and as a “center of narrative gravity” (Dennett 2013, 333–340). Dennett writes that “all of the work done by the imagined homunculus in the Cartesian theater has to be broken up and distributed around (in space and time) to lesser agencies in the brain” (Dennett 2017, 354). I also include Douglas Hofstadter’s (2007) notion of the “I” as a “strange loop.” Hofstadter writes: “An ‘I’ loop, like an audio feedback loop, is an abstraction—but an abstraction that seems immensely real, almost physically palpable…” (Hofstadter 2007, 180). According to Hofstadter “the ‘I’ [is] a hallucination perceived by a hallucination,” or “a hallucination hallucinated by a hallucination” (Hofstadter 2007, 293). There is also Thomas Metzinger (2004), who in his book suggestively titled Being No One: The Self-Model Theory of Subjectivity tells us that

no such things as selves exist in the world: Nobody ever was or had a self. All that ever existed were self-models that could not be recognized as self-models. The phenomenal self is not a thing but a process—and
the subjective experience of being someone emerges if a conscious information processing system operates under a transparent model. (Metzinger 2004, 1)

Others who might be mentioned include Susan Blackmore (2017, 67–82); illusionists (Frankish 2017); Daniel Wegner (2002); Martha Farah and Andrea Heberlein (2010); and to some extent, Galen Strawson, whose “Transience View of the self” holds that “there are many short-lived or transient selves, if any at all” (Strawson 2009, 9). For our purposes, though, our guide will be the philosopher Neil Levy (2018).

Levy notes how some existentialists undercut the belief in objective values. For some this lent a certain bleakness to the world. However, Levy states that the view expressed by contemporary cognitive scientists is yet bleaker (Levy 2018, 111). Levy writes:

But existentialists remained confident that there was someone, an agent, who could be the locus of the choice we each confront. Contemporary cognitive science shakes our faith even in the existence of the agent. Instead, it provides evidence that seems to indicate that there is no one to choose values; rather, each of us is a motley of different mechanisms and processes, each of which lack the intelligence to confront big existential questions and each pulling in a different direction. (Levy 2018, 111).

Instead, we are each of us multiply divided minds, and much of our mind is opaque to introspection. These facts spell trouble for the claim that we choose our values freely; cognitive science threatens to dissolve the self and thereby the very agent who was supposed to do the choosing. (Levy 2018, 114).

Cognitive science lends support to the modularity of the mind and the view that these modules are “functionally discrete.” There is no CEO; rather there are only “unintelligent mechanisms” (Levy 2018, 115). Levy writes that “there may be a genuine case for thinking of behavior as driven by temporary or persisting coalitions of processes” (Levy 2018, 117).

---

6. Levy contrasts the view of the mind that his is presenting with the account of the mind in folk psychology. He writes: “The account of the mind as modular is deeply at odds with our folk psychological conception of ourselves as unified beings, delegating top-down to constitutive mechanisms. Instead, it reveals each of us as a multiplicity; more community than a single organism. Worse, the community is fractured: our modules
Levy adduces varieties of evidence to support this thesis of the modularity of the mind. This includes double dissociations between processes that are suggestive of brain localization for specialized functions. Some of the evidence comes from the study of anomalous conditions. As an example, he discusses Capgras syndrome and prosopagnosia. Capgras syndrome is a condition in which, due to brain injury, a person fails to register the expected emotional response even though they recognize a face. The double dissociation is established with cases of prosopagnosia, in which the emotional reaction is intact, but the individual is unable to recognize faces (Levy 2018, 116).

As other evidence of “fractionation” and of modules with competing goals and values, Levy cites disinhibition displayed by dementia patients, individuals who exhibit anarchic hand syndrome, and cases of weakness of will (Levy 2018, 117).

Nevertheless, Levy does not totally abandon the notion of a self, but it is “an achievement and not a given” (Levy 2018, 121). Levy writes:

There is nevertheless a case for thinking that something like a self can be constituted out of this motley, a self with goals that it may pursue and which it may choose. We are limited and constrained beings, but we can impose a degree of unity on ourselves and a purpose on our lives. (Levy 2018, 121)

The modules may become functionally integrated as to form a single system that can be identified with the self (Levy 2018, 121–122). The self is “the entire collection of mechanisms” and is therefore not to be identified with consciousness acting as a CEO (Levy 2018, 117–118). In this view, the unity of the self is always a fragile accomplishment.

Clearly this view presents a complex picture. The athlete has competing modules that, to a greater or lesser extent, may cohere with one another. The same holds true for the coach. Within each athlete and coach there will be competing forces—a kind of internal athletic competition, requiring internal “team reasoning” (see Papineau 2017, 131–144). Somehow, this must all be welded into a cohesive team effort. Given this view, it is remarkable that we find consistency in athletes and coaches, and in their interactions, to the extent that we do.

This view has explanatory power. It accounts for much that transpires beneath the level of consciousness. It helps explain the double-mindedness of athletes and coaches
alike. It complicates the notion of loyalty in sport. And to the extent that we tie praise and blame to transparent choices made by conscious agents, the grounds and targets for these ascriptions are blurred.

There are various ways of looking at this view, depending on the degree of agency that we ascribe to modular processes and the degree to which they can be unified. Is there a self? Are there many selves? One’s response may hinge on where one sets the threshold for agency.

Our third answer to the question Who’s on First? is perhaps in some ways less ambiguous. Nevertheless, it poses its own puzzles.

III Many Are on First

There are remarkable, rare cases where it seems that multiple agents, rather than subpersonal modules, inhabit one body. The famous work by neuroscientists Roger Sperry and Michael S. Gazzaniga on so-called “split-brain” patients poses the issue in one way (See Gazzaniga 2016). The work involved the study of patients with intractable epilepsy whose condition was treated by severing the corpus callosum, a neural tract that serves as a major communications thoroughfare between the two hemispheres of the brain. By severing the corpus callosum, doctors are able to stop the spread of abnormal electrical activity from one hemisphere of the brain to the other hemisphere, and thus they are able to attenuate seizures. The now famous tests on postoperative patients led to questions as to whether post-surgery there were two separate consciousnesses at work (Gazzaniga 2011, 44–73). Furthermore, might there have been two separate consciousnesses—one mute, since language seems often to be centered in the left hemisphere of the brain—prior to severing the corpus callosum?8

Another condition involving multiplicity is dissociative identity disorder (DID), a controversial diagnosis that was formerly referred to as “multiple personality disorder.” In this case, distinct alters are housed by a host. Often, the different alters are seemingly unaware of the existence of one another. The condition is thought to arise out of experiences of trauma, and as such it bears resemblance to PTSD. The traumatic

7. Again, see Schechter’s recent (2018) important work on split-brain phenomena.
8. While I recall reading this point somewhere, I cannot attribute it with certainty.
experiences can be so difficult to bear that distinct identities arise to help the individual cope with their lives (see Walker 2008).  

Many will remember Herschel Walker as the former outstanding running back for the University of Georgia and, later, the NFL. He recently lost his bid to become a Senator from the state of Georgia. As an adult he was diagnosed with DID. He claims to have had as many as 12 alters, some of which were aware of the presence of other alters. Walker traces their possible origin to traumatic childhood experiences, involving verbal and physical abuse. The alters have different personalities. One alter is aggressive, and another one is consoling. Walker acknowledges that these different identities have been helpful to him (Walker 2008).

Here we have the competition between modules mirrored but at a different level. Once again, the unified self, is a task, mediated by therapy, which can bring about a convergence of alters. The condition may go undetected, and it poses a significant challenge for a coach who may not be aware of this condition. Recognition of the condition is complicated by the fact that at some level it resonates with each of us. We are all on a spectrum. But what may be mistakenly taken as mere inconsistency or moodiness may in some cases be a manifestation of a more profound reality. Which alter of the athlete has shown up for a practice or game? And for that matter, which alter of the coach?

Yet a third phenomenon involving multiplicities is presented in the case of “plurals.”

Insofar as we are dealing with multiplicity, plurals share a similarity with those diagnosed with DID. Plurals, in particular, claim that multiple persons inhabit one body. This is in contrast to singlets, who claim that one person inhabits their body. The experiences of many plurals don’t match the diagnostic criteria for DID. First, in the case of plurals, the different entities may communicate with one another. Second, plurals may not experience the plurality as a pathological condition. Instead, they seek respect and understanding (Schechter 2020). The philosopher Elizabeth Schechter writes: “Plurals don’t just feel as though they are psychologically multiple – they believe that they are. And they take each of these psychological beings, inhabiting one shared body, to be a full person” (Schechter 2020). Schechter adds, “a plural human being isn’t a person, but a co-embodied group of people” (Schechter 2020) Plurals do not mean for their claims to

---

9. There are numerous credible Websites that discuss this condition. See, for example, the National Institute of Health’s “Dissociative Identity Disorder,” National Library of Medicine, National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/books/NBK568768/, accessed June 26 2023.
be taken metaphorically. Nor do they merely mean that they have different sides. Rather, multiple persons inhabit one body (Schechter 2020).

Schechter suggests that plurals can teach us something about respecting identity. In the case of plurals, showing respect might be manifested in acceptance, which she does not conflate with belief (Schechter 2020).

Taken at face value, plurals’ claims have deep metaphysical and ethical implications. But they also have practical implications. Consider sport once again for illustrative examples. How do plurals execute split-second decisions on the playing field? What happens when there is disagreement? Which person(s) merit(s) praise or blame? How free is any person? To whom should a coach address instruction? Perhaps future research will illuminate these and other practical issues.

IV Conclusion

So, who’s on first? The array of responses by philosophers and cognitive scientists and personal testimonies present a complex picture perhaps no less dizzying than the comedy routine of Abbott and Costello. We are presented with a range of options. Each of us is a single, more or less unified entity. Or, we are rather a co-op, whose members are either known or unknown to one another, and who compete with one another. Or, perhaps each of us is (paradoxically as it sounds) no one, at least not in the traditional sense. To paraphrase from the old TV show, “To Tell the Truth,” will the real person on first please stand up?

We are also left with practical questions. How does this all work in the real world, and what can we do about it? Should we seek to be a one in the many—a unified self that exemplifies wholeheartedness? Is that even possible? Whichever view of the self that we adopt, or seek to realize, it will present challenges for understanding and negotiating the world of sport. And no one view comfortably covers all of the data.

As we shift from one view of the self to another, everything changes. And yet, everything remains the same. For, whichever view is correct, it is, though perhaps unknown to us, a reflection of the world we actually live in. And that is remarkable in itself, whoever is on first.
References


Fry


Persons, Person Stages, Adaptive Preferences, and Historical Wrongs

Mark E. Greene
University of Delaware

Biography
A former veterinary surgeon, Mark Greene completed master’s degrees in philosophy at the Universities of Hull and Bristol in the U.K., then his Ph.D. at Stanford University. After a Greenwall Fellowship at the Johns Hopkins Berman Institute of Bioethics, he joined the Philosophy Department at the University of Delaware, where he is an Associate Professor. He has a range of research interests in both applied and theoretical ethics. Many of these projects relate, in one way or another, to questions of our duties to future people.

Acknowledgments
I would like to thank Judy, Sheldon, and Talia Greene for their support and insights in discussing these issues. Thanks also to Simon Cushing and participants at the June 2023 ‘Persons’ conference at University of Michigan, Flint.

Publication Details

Citation
Abstract
Let’s say that an act requires Person-Affecting Justification if and only if some alternative would have been better for someone. So, Lucifer breaking Xavier’s back requires Person-Affecting Justification because the alternative would have been better for Xavier. But the story continues: While Lucifer evades justice, Xavier moves on and founds a school for gifted children. Xavier’s deepest values become identified with the school and its community. When authorities catch Lucifer, he claims no Person-Affecting Justification is needed: because the attack set Xavier on his life’s path, it’s no longer true that the alternative would have been better by the standard of what Xavier now values most. An unappealingly paternalistic way to hold Lucifer to account is to discount Xavier’s preferences as merely adaptive. Instead, I propose understanding the persons of Person-Affecting Justification to be not persons but person stages. This allows us to hold Lucifer to account without having to discount Xavier’s actual preferences, and has interesting implications for compensatory justice, including making sense of reparations for historical wrongs.

Keywords
Persons, Person Stages, Adaptive Preferences, Person-Affecting Ethics, Compensation, Reparations, Historical Injustice

A Six-Word Story
Let’s start with a minimalist story. (Taking inspiration from, Thomas 1966.)

Lucifer and Xavier: Lucifer breaks Xavier’s back, paralyzing him.

My hope is to make progress in understanding the ethical evaluation of this story in person-affecting terms. The first step will be to say something obvious about why Lucifer’s attack demands justification. We will then play the story out over a few more decades and see how developments complicate that obvious evaluation by suggesting a way for Lucifer to abdicate responsibility for his act. Since the problem arises from how Xavier comes to place greatest value in the actual projects and relationships he develops over time, one solution is to discount suspect preferences as merely adaptive. I agree with Elizabeth Barnes, who has argued that there should be a high bar to discounting people’s actual preferences. I propose that a better response is to take a finer-grained approach by
treated the objects of person-affecting evaluation to be not persons but person stages. I will show how this keeps Lucifer on the hook but, appropriately, loosens the connections between owing compensation and being the same person who committed the wrong, and between being owed compensation and being the same person who suffered the wrong. These lessons have interesting implications for thinking about reparations for historical wrongs. A person stage approach gives a straightforward response to the thought that reparations don’t make sense long after all those directly involved are dead, and it reveals surprisingly close connections between the case for reparations and ordinary thinking about compensatory justice.

**Person-Affecting Evaluation of Lucifer and Xavier**

Our six-word story doesn’t show Lucifer in a favorable light, but it’s too thin on context for definitive condemnation: Maybe he was acting in justified self-defense? Even so, at the very least, we can say that Lucifer’s act demands justification because of the harm to Xavier. To capture this, I propose the following principle concerning the person-affecting evaluation of acts:

**Person-Affecting Justification**: An act requires Person-Affecting Justification if and only if an available alternative would have been better for someone.

By this standard, Lucifer’s act requires justification because the available alternative of not breaking Xavier’s back would have been better for someone.

Before we complicate matters by continuing the story, there are three points worth flagging. The first is that Person-Affecting Justification is person-affecting because it concerns the justification of person-affecting acts, not because the justification must be given in person-affecting terms. Justifications can appeal to person-affecting considerations, such as breaking Xavier’s back being necessary to avert even greater harm, but Person-Affecting Justification doesn’t preclude other kinds of justification. Perhaps, never mind why, Lucifer has promised to drop a big rock at a specific time and place but, when the time comes, Xavier happens to be in harm’s way. If Lucifer has been reading way too much Kant, he might think that an absolute injunction against promise-breaking justifies going ahead despite not doing so being better for Xavier. The second flaggable point is that Person-Affecting Justification is a maximizing principle: according to Person-Affecting Justification, if Lucifer had refrained from attacking, this would still have
required justification if Lucifer could have done even better for Xavier by also giving him his lunch-money. The third point, rather obviously given the second, is that giving the required justification needn’t be all that hard: “It’s my lunch-money” will probably suffice to justify Lucifer’s decision not to bless Xavier with a cash gift.

**The Story Continues**

We are now ready to expand beyond our story’s opening six words. Don’t worry, it’s still plenty short:

**Lucifer and Xavier**: Lucifer breaks Xavier’s back, paralyzing him. In the immediate aftermath of his attack, Lucifer escapes. The following months are hard for Xavier but, after a period of understandable wallowing, he re-groups and sees how shallow and unsatisfied he’d become in the gadabout lifestyle he’s now had to abandon. Rediscovering his passion for education, he founds a school for gifted children. Working closely with the school’s dedicated staff, Xavier forges deep friendships and shares great pride in nurturing the children’s remarkable talents.

When authorities catch up with Lucifer, he shrugs off demands for a justification of the person-affecting impact of his attack on Xavier. In the time since the attack, he reasons, it has ceased to be true that the alternative would have been better for Xavier. In fact, it would have been much worse by the measure of what Xavier values most in his life: the particular relationships and projects that would otherwise never have existed.

The problem of adaptive preferences, a version of which Lucifer has raised, must give us pause. There’s the respect that makes calamity of so short life, for time will crystallize our general hopes for friendship into our specific friends. So for Xavier to wish away the wrong, is with that wish to wish away his friendships and his life’s accomplishments. That’s not a trade he’s likely to embrace. But what are these adaptive preferences, and how do they make trouble for us here?
The Problem of Adaptive Preferences

The problem of adaptive preferences, as I shall understand it, boils down to a tension between two inclinations. One is that Humean inclination not to get into fights about whether a preference is rational (Hume 1888, 416), but to accept the diversity of what matters to different people. This suggests reading ‘better for someone’ in Person-Affecting Justification, by the standard of what that ‘someone’ prefers. Pulling the other way is the Stockholm inclination to discount preferences when we deem their provenance to be somehow defective. In the eponymous Stockholm syndrome, for example, captives’ preferences seem to come into alignment with those of their captors. The Stockholm inclination is to brand such dubious preferences as ‘merely adaptive’ and to discount them in evaluations of what’s ‘really’ better for the person whose preferences have undergone such problematic adaptation.

My first inclination, on hearing about Lucifer’s treatment of Xavier, was to say that Lucifer owed a justification of his act because, in accordance with Person-Affecting Justification, the alternative of not attacking would have been better for Xavier. Only when Lucifer pressed the Humean inclination, did we see how Xavier’s commitment to his actual friends and projects complicates the judgment that an uninjured alternative would have been better for Xavier. This is awkward for those of us who’d still like to wag a stern finger at Lucifer and demand a justification on Xavier’s behalf. To press a Stockholm strategy against Lucifer we need to find a standard by which the uninjured alternative would have been better for Xavier, and we need to justify prioritizing that standard over Xavier’s considered preferences for his actual friendships and projects.

A good candidate for an alternative ‘better for Xavier’ standard is that of his total, lifetime wellbeing. Other candidate standards include capabilities (Nussbaum 2003; Sen 1985), and ‘objective’ lists (Shiffrin 2012; Harman 2009). I will stick to lifetime wellbeing here, but the points I will raise also apply to other candidate standards. We’ve noted the value that Xavier places in his school and those associated with it, but he did not need to be injured to develop friendships and projects. Uninjured, the particular friendships and projects would most likely have been different, but there’s no reason to think that Xavier would have valued them any less than he does his actual commitments. As far as having valued friendships and projects then, let’ suppose it’s a wash: the particulars would differ, but Xavier would have had friendships and projects either way. But whatever the particulars of one’s friendships and projects, wouldn’t being ambulatory add something to one’s enjoyment of them? Though the link between disability and wellbeing is not as
clear as many people, especially non-disabled people, suppose (Moller 2011), let’s assume that becoming paralyzed does take a bite out of Xavier’s lifetime wellbeing. Therefore, if we read ‘better for’ in terms of lifetime wellbeing, Lucifer does owe Xavier a justification according to Person-Affecting Justification.

It is a commonplace that people’s preferences can be at odds with their best interests: I’m even told that there are people with quite settled preferences for beer and pizza over broccoli and exercise. But are we justified in prioritizing our evaluation of Xavier’s lifetime wellbeing over his considered preference for his actual friendships and projects? Elizabeth Barnes has argued, persuasively, that it is very difficult to give substantive and non-question-begging criteria for setting aside someone’s actual preferences when evaluating what’s better for them (Barnes 2009). This leads her to caution that, “in establishing warrant for diagnosing adaptive preference behavior, the bar should be set high” (Barnes 2009, 9).

Barnes is right to set a high bar for dismissing someone’s preferences as adaptive, and the case for diagnosing Xavier as suffering a case of adaptive preferences does not even clear a low bar. The issue is that much of what we value most deeply tends not to be reliably tied to wellbeing (or capabilities, or normal function, or any other ‘objective’ standard). We look forward to developing fulfilling friendships and life plans, but when those unspecified hopes are actualized in specific people and projects, abstract valuing of friendship is eclipsed by specific commitments to actual friends, and a general hope for worthwhile work shifts to a concrete commitment to the specific projects in which we become invested. We might plan for children in the abstract, but we love them in particular. Admittedly, on very rare occasions in their children’s mid-teens, some parents find themselves wondering if their progeny couldn’t be slightly improved. Even so, if some supervillain with a time machine tells a parent he’ll go back and switch the universe to an alternate possibility with a different and ‘better’ child, that’s a threat, not a promise.

Lucifer Does Have Some Explaining to Do

Lucifer has undermined our demand for Person-Affecting Justification by holding attention on Xavier’s retrospective evaluation of his life overall. Looking forward, generalized hopes for future friendships are interchangeable regarding the particular people who end up fulfilling those hopes. But looking back, commitments to our actual friends are not fungible which is why, in retrospect, Xavier would not give up the life he’s built, not even for the promise of a bigger bucket of wellbeing. This attitude can
be healthy as, for example, when running across your middle-school bully at a 25 year reunion. Focusing on the greater significance of particular people and projects you now care about over the sunk costs of welfare lost to past shenanigans, makes it much easier to shrug and move on. For someone like me, a father steeped in the non-identity problem and constantly aware of how little it takes to change who comes into existence, it is impossible to sustain whole-hearted regret for any wrongs suffered or mistakes made prior to my child’s conception: without them, he almost certainly would never have been. However, none of this blunts the feeling that Lucifer shouldn’t be allowed to evade demands for justification just by hiding out for a while.

Instead of overriding Xavier’s values, I propose holding Lucifer to account by adopting a finer-grained interpretation of Person-Affecting Justification. Specifically, we should make person-affecting evaluations in terms of person stages or persons-at-times, not in terms of whole persons across time. If we understand Person-Affecting Justification in terms of person stages, then Lucifer does need to justify himself. In the aftermath of Lucifer’s attack, many of Young Xavier’s then current friendships and projects were derailed by his injury along with many of his more nebulous hopes for the future. Though Old Xavier prefers his actual life, Young Xavier did not: the alternative to Lucifer’s attack would have been much better for someone, namely Young Xavier.

**Not ad Hockery?**

It’s only fair to admit that there is an answer I want here: I want to wag a stern finger at Lucifer, and I want to do so on Xavier-affecting grounds. In these circumstances, best practice bids us beware of *ad hockery*. You should be suspicious that the conclusion I want to sell is loaded in a cart that’s leading my argumentative horse to market. Are person stages just a convenient trick to cobble together a way of saying that Lucifer must justify himself, or can a finer-grained approach be independently motivated? I offer no prize for guessing that I incline to answer ‘no’ and ‘yes’ respectively, but I do offer three reassuring considerations as additional motivation for basing Person-Affecting Justification on person stages.

The first reassurance is that we routinely tradeoff between person stages, both within and between persons. My pension savings sacrifice current jollies to finance bingo nights for future stages of me, and trusts trade present pleasure for the benefit of future stages of other people. Current person stages apply themselves to grueling, logic problem-sets so that their own future stages will enjoy the wealth that flows to well-credentialed
members of the philosophy profession. Having cashed in, they apply themselves to grading problem-sets, so that future stages of other people can be similarly blessed.

The second comfort is that both owing and being owed compensation occurs, almost inevitably, between person stages. Once authorities catch up with Lucifer, the commonsense thought is that Lucifer owes Xavier compensation. In the ordinary run of things, what this means is that a later stage of Lucifer will pay compensation for something done by an earlier stage, and a later stage of Xavier will receive compensation for harm inflicted on an earlier stage.

The third solace is that zooming in on person stages just re-states ordinary person-affecting evaluation with a bit more detail. Nothing is added and nothing is taken away, it’s just a matter of noticing what was there all along. My talk of person stages adds no metaphysical baggage about person stages being more fundamental than persons (Lewis 1976). People exist at times and person stages are just persons-at-times. We can distinguish specific things that are true of Young Xavier from things that are true of Old Xavier, and that’s all we need for person-affecting evaluation. Thinking in terms of person stages doesn’t take anything away either. Any truths about whole persons supervene on truths about persons-at-times, so nothing goes missing if we use the finer grain. On a person stage reading of Person-Affecting Justification, an act will still require justification if an alternative would have been better overall for some person across time, it’s just that we will be noting how this is true because of how things could have been better for that person at various stages of their life.

Limitations of Person Stages

We have seen how attending to the finer grain of person stages broadens the range of person-affecting acts that Person-Affecting Justification identifies as demanding justification, but is it still too narrow? Derek Parfit has been as influential as anyone in endorsing intuitions along the lines of something being wrong with conceiving a disabled child now when you could wait a month and conceive a different child without a disability (Parfit 1982, 118). Versions of this intuition are widely shared and have even been enshrined in British law with a prohibition on using genetic screening to select for disability (Human Fertilization and Embryology Act 2008, Section 14, Subsection 4). These non-identity cases escape Person-Affecting Justification because the alternative of never existing is not better for the disabled child. With Steven Augello, I have argued that many of these intuitions against creating disabled people should be dropped, as they...
are incompatible with even stronger commitments to reproductive autonomy (Greene and Augello 2011). Even so, I acknowledge that there is still a non-identity problem, and I do not see that a person stage approach to person-affecting evaluation suggests an easy solution to it. On the other hand, I don’t see that this puts person stages at any disadvantage to alternative approaches to person-affecting evaluation.

**Review Thus Far**

The problem posed by Lucifer and Xavier was that, in the decades following the attack, Xavier developed deep commitments to specific people and projects. The value Xavier places in his actual friends and projects is not fungible, making him rather keep those friends he has than fly to others that he knows not of, even if the alternative friendships and projects of an uninjured life would have yielded more wellbeing overall. Assuming we don’t want to let Lucifer off the hook, we toyed with the idea of discounting Xavier’s actual preferences as merely adaptive, and substituting an evaluation based on welfare, or capabilities, or some other ‘objective’ standard, even though that is sharply at odds with Xavier’s deepest values. Imperiously brushing aside what people most care about is deeply unappealing. On balance, I find the more promising alternative, which was hiding in plain sight, is to notice that even if Old Xavier endorses the actual course of his life, Young Xavier did not. By making our person-affecting evaluation in terms of person stages, Person-Affecting Justification calls upon Lucifer for a justification of his attack because the available alternative of not attacking would have been far better for Young Xavier.

**Reparations for Historical Wrongs**

Evaluating Lucifer and Xavier in terms of persons-at-times invites us to wonder how that approach might inform person-affecting evaluation in other scenarios. I will consider the issue of reparations for historical wrongs, such as slavery in the United States, as one more example of what a finer grain can reveal. There are many ethical and practical challenges tied up in this this debate, and I will not attempt to resolve them all here. Assuming that claims for reparations are, at least in part, claims for compensation, I will focus on a family of fundamental challenges grounded in the thought that, decades or centuries after an historical wrong, it doesn’t make sense to say that people who weren’t even alive to commit the wrong owe compensation to people who weren’t even alive to
be victims of it (Morris 1984). I will show how paying attention to person stages shows that reparations for historical wrongs not only make sense, but are surprisingly closely aligned with ordinary claims for compensation.

The response to the doesn’t-make-sense challenge is implicit in the point noted above, that compensation is typically paid by a later person stage for the sins of an earlier stage, and it is typically paid to a later person stage in recompense for indignities suffered by an earlier stage. Thus, to say that Old Lucifer owes Old Xavier compensation for harm inflicted by Young Lucifer on Young Xavier is just a statement of business as usual in terms of person stages. This way of stating business as usual applies, without any modification, to the payment of compensation for historical wrongs: present person stages pay compensation for harms inflicted by past stages, and present stages receive those payments for harms suffered by past stages. Since they both have the same underlying structure, if it makes sense to say that Old Lucifer owes Old Xavier compensation, then it makes exactly the same sense to say that compensation is owed for historical wrongs. On its own this doesn’t get us far, because to make sense of a claim is not to justify it. The real challenge, then, is to defend a substantive account of the kinds of links between past and present person stages that are needed to support compensation claims. This paper only gets us to the starting line of this real challenge, but I will close with some thoughts about how the way ahead might look.

Let’s start with the most obvious criterion for linking person stages, that of being stages of the same person. As a first gloss on moral common sense concerning compensation, we might say that being stages of the same person are both necessary and sufficient for both owing and being owed compensation. Thus, Old Lucifer owes compensation for Young Lucifer’s attack because they are stages of the same person, and Old Xavier is owed compensation for the harms inflicted on Young Xavier because they are stages of the same person. Conversely, we might say that no compensation is owed for historical wrongs either by or to any current person stage, because none have same-person links to stages that were either a perpetrator or a victim of those wrongs. This first gloss re-states alleged common sense about compensation in person stage terms, but does nothing to elucidate or justify what it is about being stages of the same person that carries this supposed ethical weight. We should not assume that the justification for owing compensation will work in the same way as that for being owed, so I’ll consider them separately. Because being owed compensation is the more straightforward of the two, I’ll start there.

An appealing candidate for justifying the presumption that being owed compensation is passed along same-person connections between person stages is that
people tend to have special concern for future stages of themselves. It’s true that I’m putting away retirement savings for future stages of me and not for future stages of you, but I’m also putting them away for future stages of other people who are close to me and projects that I care about. Although the familiar shorthand of ‘rational self-interest’ is easily confused with the thought that special concern for oneself is somehow rationally required, it is not. Sure, we do tend to be self-interested, but not exclusively so, and often not even primarily so. For these reasons, the scope of a standard based on connections of special concern between person stages is considerably broader than that of a standard limited to same-person connections. Given that it’s the special concern standard for which, by definition, we have special concern, it would be perverse to insist on the same-person standard.

Given the loose fit between special concern and same-person standards, it is unsurprising that there are commonsense cases in which being owed compensation seems to depart from a stages-of-the-same-person standard. For example, there’s no obvious impediment to Xavier designating someone else as the beneficiary of any compensation payment that might come his way. He could do this magnanimously by gifting his claim on any future payout to someone else, or he could do it self-interestedly by selling his claim so that he can get at least some money now. Now, suppose that Xavier dies before payment is collected. Does the claim die with him? There’s no obvious reason to think so. When Xavier transfers the interest in any future payment, it ceases to be owed to future Xavier-stages. The fact that, at some point down the line there cease to be further future Xavier-stages looks irrelevant.

What happens if Xavier dies uncompensated without having designated a beneficiary of any future payout? This is a problem that, though perhaps not explicitly solved, is one to which we have standard answers. Upon death, a person’s assets, including money they are owed, transfers to their estate and is disbursed, as best as we can figure it out, in line with their special concerns. If there is a will, this gives the best evidence we have concerning the special concerns of the dearly departed, and assets, including claims on future payments, are distributed accordingly. If there is no will, we might fall back on payouts to next of kin because they reflect the future person stages for which we presume people tend to have special concern. Absent next of kin, legally, we tend to give up at that point and return assets to the state. But this is more reflective of practical and epistemic limitations for figuring out where the deceased’s special concerns lay, than evidence that they are ethically irrelevant. People often express special concern for descendants as yet unconceived and to the communities with which they identify. I suspect that multigenerational trusts are more often established for the
benefit of a person’s own descendants than for the kids next door, and many a university development office hopes that fostering a sense of alumni community will help meet fundraising goals. In the context of this broad scope of special concern for future person stages of other people, it is easy to recognize the claim of descendants of slaves on the compensation originally owed to long-dead slaves. What of compensation owed to slaves who die childless? A reasonable presumption, it seems to me, is that members of their ongoing community will be a likely focus of their special concern going forward. And what of victims of successful genocide, lacking either descendants or an ongoing community? Things do get increasingly, empirically speculative, but perhaps a reasonable guess is that members of similarly oppressed and threatened communities are plausible loci of some level of special concern.

Turning from being owed compensation, what might we say about the links that sustain owing compensation across person stages? The most obvious first gloss on this side of things is that owing compensation is transmitted along same-person connections, because those are the connections that sustain moral responsibility. Once again, this restates common sense without elucidation or justification. And, once again, there are commonsense cases in which owing compensation departs from a same-person standard. For example, suppose that Xavier’s friend Jean had vouched for Lucifer, guaranteeing his good conduct. Having voluntarily stepped up as Lucifer’s guarantor, it’s not a stretch to say that Jean shares at least some of Lucifer’s liability for paying compensation, despite the lack of a same-person connection. Allowing for voluntary assumption of responsibility suggests relaxing the same-person standard for owing compensation, but not enough to sustain owing compensation across generations: though a son of the British Empire, I neither vouch for the good conduct of my ancestors nor do I volunteer to assume responsibility for their sins.

When thinking about Xavier being owed compensation, the payment owed is an asset to which Injured Xavier becomes entitled in the immediate aftermath of the attack. At any stage, Then Xavier may retain that asset for his own future stages or, to the extent legally and practically achievable, he is entitled to transfer that asset to such future stages of other people as may be the objects of his special concerns. Right after the attack, Xavier’s asset is Lucifer’s liability. Like assets, liabilities can be transferred, to Lucifer’s guarantors or insurers for example, but he can’t disburden himself of the liability by unilaterally gifting it to someone else. Is there some other way in which liabilities can be passed onto un-consenting future person stages, perhaps even on to future generations? We can start by thinking about how the liability gets passed on to successive stages of Lucifer, as must happen if we are to justify the commonsense claim that Lucifer still
owes Xavier compensation despite the lapse of decades since the attack. For any time at which Then Lucifer owes Then Xavier compensation, Then Lucifer either discharges that liability by paying the compensation, or he does not. If he does, we’re good. But if he keeps the money, he is holding on to an asset that is rightly Xavier’s, and transferring it to a subsequent stage of himself instead. Since the asset is rightly Xavier’s, subsequent Lucifer stages have no legitimate claim over it and they, in turn, should transfer it either to a convenient Xavier-stage, or to a stage of someone else who is a legitimate inheritor of Xavier’s claim.

A reasonable principle is that, as long as Lucifer fails to transfer the benefit where it rightly belongs, it remains ill-gotten gains to which an illegitimate recipient has no legitimate claim. Now suppose that Lucifer doesn’t keep Xavier’s money for future stages of himself, but gives it to his favorite henchperson as a discretionary bonus for exemplary villainy. Now a stage of the henchperson has control over an asset to which she has no legitimate claim and, even if she is innocent of the original sin that gave rise to Xavier’s claim on the money, it’s reasonable to think she owes it back. It’s no different than if Lucifer lifted money from Xavier’s wallet and gave it to his henchperson; the money remains Xavier’s, and the henchperson, whether she knows it or not, should give it back. This line of thinking gives a person stage description of some quite ordinary thinking about the passage of legitimate claims to ill-gotten gains between successive person stages. Just as with being owed compensation, this person stage approach makes easy sense of how owing compensation could transition to future person stages independently of a same-person standard and, importantly, independently of blame: Lucifer’s henchperson holds an asset that isn’t hers, that she does so innocently doesn’t make it any less Xavier’s.

Review and Next Steps

I have proposed a person stage approach to the person-affecting evaluation of cases like that of Lucifer and Xavier, and of historical wrongs. Other than it being a bit fiddlier, there should be no objection to putting things in terms of person stages from anyone who has any place in their ethical outlook for person-affecting considerations: the person stage approach is just a finer-grained re-description of person-affecting business as usual.

We saw how Lucifer tried to evade responsibility by exploiting a whole-Xavier evaluation in combination with Xavier’s non-fungible commitments to his actual friends and projects, and we saw why this evasion gets no traction in the finer-grain of person...
stages. Applying the lessons of Lucifer and Xavier to reparations for historical wrongs, we found that ordinary standards make sense of both owing and being owed compensation passing to future person stages, even in the absence of same-person links between those stages. This is far from an all-things-considered defense of reparations for historical wrongs, but it does show how the idea makes perfectly ordinary sense, and it places the emphasis on stating and defending criteria for how both owing and being owed compensation are transmitted from one person stage to another. The next steps, then, are to address the real challenge of elucidating those ethically relevant links between person stages. I have suggested that links of special concern for future stages can sustain chains of being owed compensation that reach further than a same-person standard. For owing compensation, I’ve suggested that unpaid compensation can be viewed as ill-gotten gains such that even innocent recipients may have a duty to repay. There are many more ethical and practical complications around the disposition of ill-gotten gains and other factors relevant to the full evaluation of compensation claims (Katz 1996), but these initial considerations suggest that a person stage approach reveals surprisingly close connections between ordinary interpersonal compensation and historical reparations.

References


Biography
Eric Kraemer received his A.B. in philosophy from Yale University and his Ph.D. in philosophy from Brown University. He taught at the University of Nebraska-Lincoln and the University of Wisconsin-La Crosse. His research areas include philosophy of mind, epistemology, and medical ethics.

Dedication
Lynne Rudder Baker (1946-2023) was Distinguished University Professor of Philosophy at the University of Massachusetts-Amherst. Baker was the author of 5 books and numerous articles. This paper is dedicated to her memory.

Publication Details

Citation
Abstract
Persons are natural beings, but they probably do not constitute a natural kind. So how should a philosophical naturalist conceive of them? This article begins by critiquing different views on the nature of persons put forward by animalist, Eric T. Olson, medical ethicist, Mary Anne Warren, and common-sense metaphysician, Lynne Rudder Baker. Baker’s views are of primary interest, as she initially offered a naturalist-friendly account of persons which later morphed into an anti-naturalist position. A naturalist account is proposed based on organisms having the properties of consciousness and intentionality that borrows from and modifies all three earlier views. The proposed account identifies three different kinds of persons. Kinds of persons are differentiated from each other by their serious capacity for one of three different forms of intentionality. The three different forms of intentionality identified are basic intentionality, enriched (or social) intentionality, and the intentionality required for a fully developed first-person perspective. Limitations of the proposal are explained, criticisms of the proposal are addressed and advantages of the proposal are enumerated.

Keywords
Persons, Philosophical Naturalism, Baker, First-Person Perspective, Intentionality, Warren, Animalism

Introduction

If you are reading and comprehending this discussion, then you are most surely a person. But what is it for there to be persons? What makes a person a person? Let’s call this the ‘Person Question.’ Human beings have attributed personhood to a wide variety of objects: in addition to human beings and various animals, plants, rocks and bodies of water have also been considered persons at some point in human history. Furthermore, the attribution of personhood has typically been taken to indicate the moral significance of the individual or group in question, and also to indicate limitations on how such individuals or groups are to be treated.

1. Given the current hype regarding ChatGPT, I should note that I do not consider current computer programs, however extraordinary their features, to have yet reached a level of sophistication that would counts as genuine comprehension. But, given my commitment to philosophical naturalism, to be consistent I need to remain open on this matter.
This discussion, however, approaches the Person Question from the perspective of Philosophical Naturalism, the view that the only objects that exist are physical objects and the only properties that are instantiated are either physical or ontologically-neutral properties. No attempt will be made to accommodate non-natural entities such as souls, or spirits, or to accommodate the many celebrated purely imaginary or pseudo-scientific thought experiments often invoked in philosophical discussions of persons. This question about persons is an important one for philosophical naturalists to address. The answer to the question is by no means obvious for philosophical naturalists as persons as a group do not figure as the specific objects of any current branch of natural science.

The Person Question, we should note, is different from the Personal Identity Question, what makes a person the same person over time? But, it is reasonable to assume that any attempt to answer the Personal Identity Question must presuppose having an answer to the Person Question, as the latter is more basic. It should also be granted at the outset that even if one has a convincing answer to the Person Question, there is no guarantee that the Personal Identity Question will also receive an illuminating or definitive answer.

Ever since Peter Strawson’s influential 1958 eponymous article, the concept of persons has been an important one that analytic philosophers have been confronted with having to address.² So, it is no surprise that naturalistically-inclined philosophers have found themselves in need of either accommodating the concept or explaining it away.³ But, since I take it as obvious that there are persons and since philosophical naturalism seems the most promising philosophical framework, I will focus on how one might best accommodate the concept of persons within a naturalistic framework.⁴

In this discussion my efforts will concentrate on the development of the late Lynne Rudder Baker’s philosophical views on the topic of persons. My choice of Baker is deliberate, as Baker initially developed an account of persons which she took to be broadly compatible with a weak form of naturalism (Baker 2000), but then gradually changed her mind. At first she embraced a form of what she called “quasi-naturalism” (Baker 2007), and then she explicitly rejected the adequacy of naturalism, opting instead

---

². In this discussion I will follow the convention adopted by many writers of using the term “persons” instead of the term “people” but my usage is purely stylistic. I do not recognize any significant philosophical difference between these two terms.

³. See, for example, Peter Unger (1979).

⁴. In what follows for ease of exposition I will use the terms ‘naturalism’ and ‘naturalistic’ to refer to philosophical naturalism.
for a view which she claimed was “close enough” to naturalism (Baker 2013). Thus, Baker is constructively useful for my purposes, first, in initially being sympathetic to a naturalist account of persons, and thereby laying out a helpful initial framework that naturalists can substantially adopt. And, second, her work is also critically valuable in providing both serious counter-arguments and nagging concerns that conscientious naturalists need to consider and counter.

To guide my efforts, in addition to Baker’s writings, I will also refer briefly to the pioneering work of philosophers, Eric T. Olson and Mary-Anne Warren. I will attempt to combine elements from all three thinkers into a view that none would accept. Unlike philosophers who seem concerned to tie personhood to multiple complex forms of thinking, my primary concern is to determine what minimal natural conditions, including minimal mental abilities that are required for minimal personhood. My reason for this deliberate emphasis is that it is all too easy for philosophers (1) to forget that there are many members of our own species that lack certain basic mental abilities and also (2) to ignore the option that there are members of a number of other species that share important mental features that we human beings prize.

Some Initial Comments on the Naturalization of Persons

Any attempt to say something meaningful about persons must start with an initial list of prospective candidates for personhood. Since as humans we are by nature anthropocentric, we must start with members of our own species. If we are willing to be fully anthropocentric, as many of our conspecifics still are, then we will stop at this point, claiming that human persons have a truly unique metaphysical status and are therefore special. But, there are other influences, religious, social, fictive, in addition to personal experiences with and testimonies by others about other animals that will often force many of us to be open to the notion that the class of persons is wider than the class of humans. Although, as I have already mentioned, I shall not be concerned to accommodate religious entities or science fiction entities in my efforts, I will, on the other hand argue, that naturalists have to be open to the live option that other animals are persons, too.

5. The example of Jane Goodall talking about her experiences with chimpanzees in her numerous books and documentaries is particularly moving.
What makes the question, “what are persons?” difficult to answer is that it is widely agreed that the concept of a person is a primitive concept, that is, one that cannot be further analyzed into simpler notions. Any attempt to provide an analysis of person, such as “A person is an individual with a self-concept,” seems to utilize a term, such as self, that either itself requires explication in terms of person or is a synonym of person. What one can do with respect to trying to get a deeper understanding of a primitive concept, however, is attempt to characterize it by pointing out its significant key features.

So, what is involved in naturalizing a primitive concept? Here’s a seven step procedure to follow.

1. Identify a set of typical characteristics generally regarding as being had by those individuals to whom the concepts apply.
2. Ascertain which features are also possessed by uncontroversially natural objects and which are not. (Those features also had by uncontroversially natural objects pose no threat for naturalists.)
3. With respect to those features not had by uncontroversially natural objects, determine whether and how they might be assimilated to natural properties.
4. If certain such features can be easily assimilated to natural properties, they, too, will pose no issues for naturalists.
5. Turn next to considering those features which cannot be easily assimilated to natural properties and determine whether or not they are essential to retain the concept as consistent with naturalism.
6. If a feature is not essential, then explain why it may be ignored; if a feature is essential, then adopt a revisionist strategy for these features.
7. To implement a revisionist strategy, consider which aspects of these features are most salient, have the best evidential support, and determine natural replacements for them.

For example, one might start by proposing that a person is a living organism with certain specific mental characteristics, A, B and C. To apply the above procedure schematically we would need to consider living organisms and features A, B, and C. Now it is reasonable to suppose that living organisms are naturalistically congenial, not requiring any non-natural properties or non-physical forces to explain their operations. We would then need to turn to features A, B, and C to see if a promising naturalistic account of them
can also be given. Let us suppose further that there are initially plausible naturalistic accounts already for A and B but not for C. We would then need to ask whether feature C was a significant feature of personhood. It might turn out that we could produce an argument to show that C was metaphysically dubious, and so not worth worrying about. In which case we could revise our naturalistic account of personhood to consist simply of organisms with features A and B. But suppose, on the other hand, we determined that C introduced an important feature of personhood but was not naturalistic in its present form. In this case we would need to figure out whether there was a more satisfyingly naturalistic replacement for C, call it C*, and whether C* would be adequate to cover the cases of all those individuals whom we wanted to include as persons.

The above schematic example begins with the assumption that persons are first and foremost living organisms that also have certain mental features. From the naturalist perspective, species of organisms that are persons evolved from other species of organisms that were not persons. Therefore, considering how the development from organisms who are not persons into organisms that are persons seems to be a fruitful way of providing an answer to the person question. Since starting with organisms seems to be a productive procedure for naturalists to follow, let us begin the search for an adequate naturalization of persons by considering the views of Eric T. Olson, who makes the organism the fundamental basis for personhood (Olson 1997).

**Olson’s Animalism**

In his 1997 book Olson proposes and defends a naturalist-friendly account of personal identity based on the human being as an organism. This view has come to be known as ‘animalism.’ According to Olson, our basic organismic nature makes us what we are and accounts for our continuing to exist: “On the Biological Approach, what it takes for us to survive remains the same throughout our careers: like other animals, we persist as long as our life sustaining functions remain intact” (Olson 1997, 89). He also says, “The fetus or infant becomes a person” (Olson 1997, 89). You, however on Olson’s view, are not basically a person, you are fundamentally an organism. You can also survive your psychological demise provided your organism continues otherwise to function. Olson admits that “Perhaps we cannot properly call that vegetating animal a person, since it has none of those psychological features that distinguish people from non-people (rationality, the capacity for self-consciousness, or what have you)” (Olson 1997, 17). Thus, Olson’s view seems as though it might well count as a naturalized account of
personhood. Applying our above schema, for Olson a person is an organism plus the acquired features of (A) rationality, (B) the capacity for self-consciousness or (C) “what have you.” The problem for the picky naturalist, of course, will be with respect to (C), that is, not knowing exactly which specific feature or features the acquired possession of which Olson thinks turn organisms into people. But, unlike defenders of psychological criteria for personal identity, Olson is not worried to spell out what exact criteria, from a psychological perspective, a person phase of an organism needs to have. Thus, it is reasonable to assume that he does not consider such criteria to be metaphysically controversial. His metaphysical concerns are elsewhere. In particular, he is convinced that personal identity is organismically grounded, not psychologically based. So, while Olson distinguishes between the mere organism and its person phase that your organism transitions into and out of, it is the continuity of the organism that is essential for your continuity.

Olson maintains that “I” was an organism before “I” was a person, and that “I” may biologically continue as an organism after I cease being a person. But, why should we interpret things this way? Cases in which living bodies which cease relevant psychological function seem to me to have lost what is essential to their “I”-hood. Olson’s alternative interpretation is simply odd. Here’s a quick-and-dirty linguistic argument. The word “I” functions as a subject. What seems precisely to be lacking, if we try to take Olson seriously, is how statements about organisms lacking all psychological properties can be meaningfully interpreted as referring to subjects.

While as a naturalist I completely agree on the importance of my organismic nature’s role in making me what I am, I think Olson’s approach to persons is insufficiently organismic; he fails to appreciate the remarkable biological developments that are required for persons to exist. These biological developments are what really turn organisms into persons; Olson even admits there are no persons without them. But, my comments just reflect my own intuitions, perhaps Olson would reply that he just does not share these intuitions. So what arguments do serious critics provide?

Consider Baker. Baker raises two objections to Olson’s view, one based on the possibility of replacing all of a human’s biological parts with inorganic parts (Baker 2000, 122), the second based on brain transplants (Baker 2000, 124). Baker claims that both of these scenarios are conceivable, but that Olson’s view cannot accommodate them. But, there are replies to both of these science fiction possibilities. Olson can maintain that an organism with perfectly functioning artificial parts artificially performing all of the required bodily functions of the organism is still an organism, although an artificial one. And, Olson bites the bullet on brain transplants, holding that one’s brain switching
organisms is equivalent to one’s death as a person. “So if you are a human animal, you do not go along with your cerebrum when it is transplanted; you simply lose an organ, and with it those psychological capacities that depended on that organ” (Olson 1997, 18).  

As a naturalist, my own view on this matter is that we simply do not know what the results of such a transplant might be. I can well imagine both scenarios that would support Baker’s objection as well as contrasting scenarios that would support Olson’s reply. At this point, there does not seem to be a way to decide whether Baker’s transplant objection has any real force. We should instead remember that science fiction really is fiction, and not put much confidence in what are, for now, purely fictive scenarios. So, I do not find Baker’s objections to Olson’s view terribly compelling. In fact, I am not currently aware of any blatant inconsistency in Olson’s animalist view; I just fail to find animalism to be a convincing account of persons. It would seem that a much more compelling alternative to Olson’s animalism would be to propose, not that a person is simply an organism, but rather that a person is an organism as long as they have certain requisite psychological properties. But, to move to this next step we first need to consider an important contribution to the abortion debate by another philosopher, Mary Anne Warren, whom I view as an important forerunner to Baker.

**Warren’s Thought-Experiment Approach**

It should not be surprising that one of the most important areas in which the Person Question should have been seriously raised has been in the debates over abortion over the past half century. After all, if the fetus is a person at some particular stage, then this seems to be nearly decisive regarding whether abortion is morally permissible or not. And, if the fetus is never a person, then the debate over abortion seems moot. The latter view is famously defended by Mary Anne Warren in a classic article in which Warren attacks leading views of the day (those of John Noonan and Judith Thomson) as well as offering her own innovative argument (Warren 1973).  

For my purposes, John Noonan’s view is worth considering briefly. Noonan claims that a fetus is a person because it possesses (a) “a full genetic code” and (b) the potential

6. Olson does not share the same religious commitments that we shall later see that Baker has.
7. Exceptions to be considered include a threat to the life of the mother, pregnancy as a result of rape or incest, and a dire medical diagnosis of an unavoidably and terribly painful, short and meaningless life of the future neonate
capacity for rational thought (Noonan, 134). Although Noonan was a prominent defender of the conservative religious view on abortion, Noonan’s account of personhood includes an explicitly naturalistic part, namely the genetic code, and adds an important mental element, the capacity for rational thought. If this mental element could be accounted for naturalistically, then Noonan’s account would be one that naturalists could also accept. Applying our above schema, Noonan’s analysis consists not of an organism but rather of genetic material in a developing fetus combined with a single mental property, the capacity for rational thought.

We also need to remember that Noonan claims that having a full genetic code is already a sufficient condition for also having the capacity for rational thought, which, he thinks, makes abortion at any stage of fetal development prima facie immoral. Warren deftly critiques Noonan’s claims, arguing that the most that Noonan can establish is, not that the fetus is a person at every stage in virtue of it possessing the capacity for rational thought, but, rather, the much weaker claim that the fetus has the potential to become a person in virtue of possessing this capacity, which carries no force with respect to rejecting the permissibility of abortion.\(^8\)

Warren instead proposes a novel thought experiment in which “a space traveler… lands on an unknown planet and encounters a race of beings utterly unlike any he has ever seen or heard of” (Warren 1973, 54–55). Warren suggests that, in order for the space traveler to figure out whether these beings had moral standing (i.e., were persons,) the space traveler needs to consider five characteristics she claims, “are most central to the concept of personhood” (Warren 1973, 55):

1. Consciousness (of objects and events external and/or internal to the being), and in particular the capacity to feel pain;
2. Reasoning (the developed capacity to solve new and relatively complex problems);
3. Self-motivated activity (activity which is relatively independent of either genetic or direct external control);
4. The capacity to communicate, by whatever means, messages of an indefinite variety of types, that is, not just with an indefinite number of possible contents, but on indefinitely many possible topics; [and]

---

8. For a further critique of Noonan-type views see Kraemer (1983).
5. The presence of self-concepts, and self-awareness, either individual or racial or both.

Warren’s strategy to counter the anti-abortionist’s claim that fetuses are persons is to assert that at no point in its development does a fetus possess any of these five characteristics. Since the fetus lacks all of these features, it is not a person, but only develops into a person at some later point after birth (Warren 1973, 56).

Warren’s discussion is relevant here because it introduces an interesting variety of features and also allows for some flexibility regarding the attribution of personhood. In particular Warren says:

We needn’t suppose that an entity must have all of those attributes to be properly considered a person; (1) and (2) alone may well be sufficient for personhood, and quite probably (1)–(3), if “activity” is construed so as to include the activity of reasoning. (Warren 1973, 55–56)

Warren’s thought-experiment approach is also important as it seems to allow for the possibility of two minimal kinds of persons, first those beings with consciousness and reasoning, and second those beings with consciousness, reasoning, and self-motivated activity (guided by reasoning). It also allows for other species to be persons. But, importantly, it also allows for more advanced persons, namely those also possessing communication abilities and self-concepts. And, if these features can all be provided with an adequately naturalistic account, then naturalists could well accept these different options as different kinds of persons.

There are some minor issues that might be raised for Warren’s account. First, Warren’s view is, rightly in my view, held hostage to scientific developments. So, for example, if it is determined that fetuses do possess consciousness and, especially sentience, as some have claimed, then this needs to be taken seriously. Warren can still maintain that an additional criterion would need to be discovered in fetuses for her view on abortion to be affected. Second, Warren’s view as stated above might seem to place an undue emphasis on the importance of a person’s possessing the ability to reason. But, it is not uncommon for some individuals to lose this capacity, either temporarily or indefinitely.

---

9. Charity requires that we recognize that scientific developments have certain changed over the past half century. A quick web search quickly reveals numerous sites claiming fetal sentience begins anywhere from 18 to 25 weeks.

10. Following Noonan and to some extent as well as Aristotle.
And, in such cases it also would seem odd to have to say that such an individual ceased to be a person, especially if they possessed other features on the list, such as the capacity to communicate and possession of self-concepts. So, perhaps a more charitable reading of Warren’s view would be to suggest holding that a person is an individual who possesses consciousness and any of the other four criteria. This would then result in an even more complicated account of personhood than what I have indicated above. Third, although in addition to humans Warren only explicitly considers applying her view to alien persons and to the possibility of self-conscious robots and computers, she does not consider what her space traveler would think of elephants, dolphins and octopuses when they return to Earth. Again, her view could be extended to include all of these other species using several of the criteria from her above list.

Historically, however, the most important objection to Warren’s position is that her view would permit infanticide up until the age of two. In a subsequent (Warren 1984) postscript to her original article, Warren defends against this charge by arguing there are other ways to argue for protecting neonates, such as the fact that they are deeply desired by their parents. But, those who think that very young infants who still possess significant capacities with respect to the above criteria are not going to be convinced. And, a related serious issue concerning euthanasia needs to be faced. How on Warren’s view of persons should we respond in the case of an adult human has temporarily lost all or most of the five Warren features of personhood? On Warren’s view it would seem that, as for the neonate, due to the absence of the requisite features such an individual is no longer a person. But, if Warren allows that the adult retain their personhood status during the period temporary loss from which they are expected shortly to recover, which seems the reasonable response, then it is blatantly unfair to deny the status of personhood to a fetus on the verge of shortly acquiring the same features.

**Baker’s Initial Constitution View**

Let us now turn to Baker’s Constitution View. In her book, *Persons and Bodies*, Baker presents a new way of conceiving persons which she claims to be consistent with a ‘weak materialism’ (Baker 2000, 134). A person, for Baker, is a combination of an adequately developed body plus “a capacity for” a first-person perspective (Baker 2000, 92).\(^{11}\)

---

11. It is tempting to view Baker as revising the fifth of Warren’s central criteria of personhood listed above, namely the presence of self-concepts, and making it central to personhood.
This combination of items is to be understood as a technical relation that Baker calls *constitution*, which is not at all the same thing as identity. For Baker, a person’s being *constituted* by a body and a first-person perspective means that (1) the body and the person spatially coincide but are not identical, and (2) it is possible for the body to exist without there being a person present.\(^{12}\)

As is the case with Olson’s animalism, Baker’s constitution account of personhood requires that personhood be something that is both gained at some point after the body came into being and is also something that can be lost while the body continues to exist. But, for Baker, persons only comes into being when their capacity for a first-person perspective is acquired. And this, in turn, requires that “all the structural properties required for a first-person perspective” are present, and the body is “in an environment… conducive to the development and maintenance of a first-person perspective” (Baker 2000, 92).

Before considering subsequent revisions of Baker’s view, here are three quick criticisms of Baker’s initial Constitution View. Consider, first, Donald Davidson’s infamous *Swampman* thought experiment (Davidson 1987). If an alien body, Herbert, metamorphosed out of organic gook right in front of us and then communicated with us well-enough to convince us that they had a 1\(^{st}\)-person perspective, it would seem utterly bizarre to claim that Herbert failed Baker’s theory of personhood simply because Herbert did not develop over time but arose spontaneously. As a naturalist I have sworn off such considerations; but, Baker, given her criticisms above of Olson, leaves herself wide-open to this concern. (And, she could modify her view accordingly.)

There is the further worry as to whether Baker’s constitution approach really does solve the problem of personal identity any better than other views she criticizes. It is simply not epistemically certain whether the same person over time is really being picked out by the same particular first-person perspective or whether there are two different first-person perspectives. Not only is this not the case for other people that one observes from a third person perspective, but it is also not the case from one’s own first-person perspective. It may seem to me that (a) I have the same first-person perspective this morning that I had yesterday morning, but, the skeptic will point out, I could be wrong.

\(^{12}\) The metaphysics behind Baker’s notion of constitution is murky. She motivates the notion by appealing to the difference between a statue and the piece of marble and art world that constitute it. But, it is hard to see a convincing parallel for the case of persons. The first-person perspective, although strongly influenced by language development and social interaction, seems much too organically connected to the well-functioning body. It would seem more appropriate to regard it as a property of the body.
There might in fact be (b) two different first-person perspectives involved. And Baker has no independent way to determine whether (a) or (b) is the case.

Baker is well aware of this objection. In a discussion in a subsequent book chapter on resurrection and the survival of death, to which she is committed, Baker admits:

What is needed is a criterion for sameness of first-person perspective over time…. Although I think the constitution view solves the synchronic problem of identity non-circularly…I think that, on anyone’s view, there is no informative noncircular answer to the question: In virtue of what do person P1 at t1 and person P2 at t2 have the same first-person perspective over time? It is just a primitive, unanalyzable fact that some future person is I, but there is a fact of the matter nonetheless. (Baker 2005, 385)

But, to be aware of the objection and to suggest in reply that identity of first-person perspectives over time is a brute fact should strike the reader as a rather spectacular piece of stone-walling. It seems reasonable to hope that something more can be said to help us track persons over time. If an alternative view of personhood can do a better job of handing the problem of personal identity, then that would seem to be a strong reason to prefer it over Baker’s constitution view.

Also, the spectre of the infanticide objection is lurking. A quick google search reveals that a sense of self develops in infants between the ages of two and three. And we know that brain structures are developing continuously until the age of 25. Determining whether the right structures are in place to avoid the infanticide charge might have made Baker’s view a hostage to future science. But, since she explicitly rejects reductive materialism and any promise of help that it might provide, it is not clear on Baker’s view how one could determine just when the right brain structures for the capacity for a first-person perspective will be in place. So, Baker’s early constitution view would also seem to face the same infanticide objection that haunted Warren.

---

13. First-person perspective identity might not be a brute fact due to empirical concerns. Suppose brain-conjoined twins are determined to have overlapping first-person perspectives, but consider themselves to be non-identical.

14. For example, defenders of animalism might claim to do a better job solving the problem of personal identity.
BAKER’s Initial Rejection of Naturalism

In her subsequent book, *The Metaphysics of Everyday Life*, Baker continues to defend the constitution view outlined in *Persons and Bodies*, but revises her metaphysical claims. In particular, she no longer claims that her views are broadly consistent with naturalism, but instead embraces what she calls ‘quasi-naturalism.’ Quasi-naturalism differs from thorough-going naturalism in two ways. First, epistemologically, quasi-naturalism holds that there are other sources of knowledge besides the sciences. Second, metaphysically, quasi-naturalism holds that there may be some events that do not have scientific explanations (Baker 2007, 87). What quasi-naturalism entails with respect to human persons also involves two claims. First, human persons are part of a natural world that has evolved by natural causes over eons, are natural entities, and live under the same necessity as the rest of nature (Baker 2007, 89). And, second, human persons are ontologically unique in that the coming-into-being of a new person is the coming-into-being of a new kind of entity, not merely an already existing entity’s acquiring a new property (Baker 2007, 90). In response, naturalists would readily admit that human beings are indeed remarkable. But, naturalists would contend, given that human persons are natural entities subject to natural forces that develop according to laws of nature, it seems highly unlikely that human persons are ontologically unique. What is not obvious is how one might argue for this uniqueness; and clearly, an argument is needed. Let us now consider the important additional modifications in Baker’s theory of persons introduced in the final version of Baker’s view.

Baker’s Two-Tiered Revision

In her 2013 book, *Naturalism and the First-Person Perspective* Baker refines her view further. The same revisions are nicely summarized in a subsequent article, “Making Sense of Ourselves” (Baker 2016). This article’s initial focus is to critique narrative accounts of personal identity as championed by Daniel Dennett and Myra Schechtman. One of Baker’s criticisms of Schechtman’s characterization view is that narrative accounts do not apply to infants (Baker 2016, 12). Perhaps to avoid a similar problem for herself, Baker now moderates her view, claiming that there are two different kinds of first-

15. I will here address remarks from her 2016 article, in part because it is most relevant to the 2023 CNN conference.
person perspectives, ‘rudimentary’ and ‘robust.’ The rudimentary first-person perspective consists of consciousness and intentionality, while the robust first-person perspective is the full-fledged “I*,” or “the capacity to conceive of oneself as oneself in the first-person” (Baker 2016, 14). Further, Baker claims that most humans after becoming rudimentary persons normally develop into robust persons. Baker is now able to avoid the infanticide objection by claiming that infants possess rudimentary first-person perspectives, which subsequently usually becomes robust.

An important consequence of this distinction is that Baker admits she must now grant that many mammals—her favorites are lions—namely those possessing consciousness and intentionality, also become constituted as rudimentary persons when their bodies develop structures which can sustain consciousness and intentionality. Still, Baker insists, a huge metaphysical distinction still exists between humans and other animals: humans have their first-person perspective essentially while nonhuman animals only have first-person perspectives contingently. Again, as with Baker’s earlier claims regarding the uniqueness of human persons, I cannot find any non-contentious reason to accept this claim; it just seems to be dogma. Further, her distinction between two classes of persons raises the worry as to how we should treat those humans who, because of bad biological or environmental luck, never become constituted by bodies with structures supporting the robust version of the first-person perspective. In addition the question as to how to treat animals with a rudimentary first-person perspective also arises.

Further, once one introduces two kinds of first-person perspective, it is then tempting to ask whether there are other kinds of first-person perspective. Might there be a moderate first-person perspective in between rudimentary and robust, consisting perhaps of having a weaker sort of first-person perspective?\textsuperscript{16} The worry is that introducing a variety of first-person perspectives makes it difficult to claim that one specific level or specific combination of mental traits is essential for personhood. The important moral we should again draw, as we have already observed in examining Warren’s view, is that any adequate account of personhood needs to be suitable flexible on this point, allowing that there seem to be very different kinds of people with very different mental endowments.

At this point, after considering and critiquing the views of Olson, Warren and Baker, I will attempt to assemble the insights I have gained into a naturalized account of persons. Unlike the previous accounts so far discussed, there are no relatively simple guiding principles, except perhaps those of avoiding bias in favor exclusively of

\textsuperscript{16} Or, might there be an even more advanced kind of first-person perspective for Aristotelian super-stars who excel at the highest forms of thought, such as contemplation?
Kraemer’s Non-constituted Alternative Account of Persons

I begin my account by first taking stock and indicating what features so far discussed I agree with and what challenges remain. I think Baker is on the right track in terms of thinking of persons as being associated with individual physical bodies. While I cannot follow Olson’s animalism, I find Olson’s organismal assumption, the idea that persons are probably a heterogeneous class of organisms with certain mental abilities, also to be convincing. The hard part involves trying to specify which minimal mental ability or which set of such abilities are required. It seems reasonable to suppose that natural persons encompass a range, many features and dimensions of which have already been mentioned.

As a naturalist, I also urge that, while we may be very clear about certain additional kinds of organisms that we want to include (octopuses)\(^\text{17}\) and which to continue to exclude (slime molds), there are others that at this time we are not at all sure about. My proposals are tentative, like many areas of investigation in science. There are certain claims that we can justifiably advance now based upon our current scientific understanding. But, we need to be humble and admit that only future science may be able to give us more definitive insight as to not only the range of individuals that should be considered persons, but also with respect to the re-identification of specific individual persons over time.

Warren’s bold attempt at categorizing persons is indeed helpful. And, it even seems plausible if we interpret her view as holding that organisms possessing consciousness and at last one other of the right elements should be considered persons. Our discussion so far has made it clear that being a naturalist regarding persons is going to be a messy proposition for the foreseeable future. There will be a variety of different sorts of persons, some very rudimentary indeed, some moderately developed, some fully developed, and some quite spiffy. And there will be hard cases where it just will not be clear what to say.

One further point. In her 2013 book Baker raises an important general point about persons, one hinted by other writers, namely the importance of language, which entails language communities. She says: “Persons are not solitary selves. They require language

\(^{17}\) Any reader doubting the personhood of octopuses is recommended to read Peter Godfrey-Smith [2017].
communities” (Baker 2013, 140). From a naturalist perspective, this claim should be taken seriously. Many species often considered as likely candidates for personhood seem also to have some form of communication, and live in communities. This leads me to introduce the notion of *enhanced* intentionality. Intentionality refers to the directedness of our thinking on objects, including some which may not exist.\(^\text{18}\) By enhanced intentionality I mean to indicate possessing sufficiently complicated mental structures that enable one to recognize individuals in a community, including recognizing oneself as an individual in that community. I am understanding enhanced intentionality as a mental capacity in between the perhaps basic intentionality of lions and the full-fledged I* ability to think of oneself as a self.

As the reader has already been warned, my view on persons is a rather messy one. Messy views are unfortunate because clear and simple philosophical views tend to be more successful at garnering attention and followers: they are easier to grasp, remember, apply, critique, revise, refute, etc. But, if we fairly consider all of the various animals that might be considered for some form of personhood as well as the corresponding different kinds of members of our own species that we would want to include as persons, then it becomes clear that, unlike some of the simpler answers that have been considered so far, the Person Question deserves a nuanced and multiple-level response. First, we would do well to consider distinguishing between at least three kinds of persons to be specified below. Second, on this view it is incumbent for us to be more humble about what the actual status of various animals actually is, and grant that we might turn out to be wildly incorrect in our current assessments.\(^\text{19}\) And, third, as a result of trying to naturalize persons we may need to admit that, although there are coherent and defensible accounts of personhood, the philosopher’s problem of personal identity over time may be one for which only a roughly approximate answer can ever be provided. We may be forced to admit that there may simply be too many obstacles to ever providing a definitive answer to all of the traditional philosophical questions that have been raised.

Here, then, are the six parts of my view of persons.

[1] *Basic* persons are organisms (living organic bodies) possessing consciousness and what I will call *basic intentionality*. By basic intentionality I include having some propositional attitudes about objects in the world, being capable of being fooled with

\(^{18}\) See Chisholm 1956, 125

\(^{19}\) For a discussion of moral issues at stake in determining the nature of animal thinking, see Kraemer (2006).
respect to those attitudes by appearances, and, further, being capable of engaging, or at least intending to engage in intentional actions based upon those attitudes.

With this requirement I intend to accommodate the initial granting of personhood to non-human animals. The details of who would be included in this category remain to be determined. Certainly, neonate humans quickly develop to this stage.

[2] Social persons are organisms that possess consciousness and what I call enhanced intentionality. If we are lucky, future science will reveal when these mental possessions are acquired as well as the extent of those that have them. In addition to apparently including a number of social animal groups, this category also includes certain human beings who face serious mental challenges.

[3] Full-blown 1st-person perspectives are indeed special, but material in nature. I think that they are constructed somehow out of consciousness and intentionality. They are remarkable and important but not ontologically unique and are not necessary for being a person.

[4] I am not especially sanguine about solving the problem of continued existence, especially not by appealing to first-person perspectives, but think a naturalist approach that looks for physical evidence in the organism is the only one with any likelihood of even moderate success. I have more to say about this below.

[5] I do not worry about constructing a theory of persons that accommodates deities, Martians, artificial devices, brains-in-vats, transplanted brains or brain-parts, or individuals who have been ‘teletransported.’ If and when good impartial evidence of such things really existing is forthcoming, there will be plenty of opportunity to investigate and revise.

[6] I think we will need a separate category of seriously potential persons to cover certain individuals of our own and many other species. This seems to be the only honest way for humans to avoid the infanticide problem and also to help provide much needed guidance in connection with the appropriate way to respect organisms at the end of their biological lives. I have no proposals to offer, but I remain convinced that scientific investigation may be able to help us figure out some reasonable answers. Let us now turn to considering some objections.

20. Teletransportation was popularized in the Star Trek television series, and famously used by Derek Parfit (1986).
Baker’s Arguments Against Naturalizing the First-Person Perspective

In her last book Baker goes on the attack against Naturalism. She contends that it is a flawed view for not being able to account for the first-person perspective. She argues as follows (Baker 2013, 123):

1. There are first-person properties that are neither eliminable nor reducible.
2. Any property that is neither eliminable nor reducible belongs in the ontology.
3. Thus, first-person properties belong in the ontology.
4. If first-person-properties belong in the ontology, then ontological naturalism is false.
5. Thus, ontological naturalism is false.

As an ontological naturalist, I must reject the argument’s first premise. But, if I do not think that first-person properties are eliminable—and I do not—and if I do not have a handy-dandy reduction up my sleeve, then how can I reasonably justify rejecting that premise?

I admit that not having a reductive strategy ready-to-hand is a bit embarrassing, but so then is lacking a plausible reductive account for consciousness and for intentionality. And, that is, oddly enough, the key to my response to Baker. After all, Baker is not worried about the natural status of rudimentary first-person perspectives of human infants and other nonhuman animals, which are not ontologically special. And, she grants that they possess consciousness and intentionality. I think it reasonable to suppose that a robust first-person perspective is somehow composed of elements of both consciousness and intentionality. If we can at least provisionally grant that consciousness and intentionality are within the scope of ontological naturalism, this suggests the beginning of a strategy to deal with first-person perspective properties as well. While Baker would insist that having a first-person perspective is very different from what I have referred to as enhanced intentionality, I am confident that further investigation into what actually happens in us when our thoughts are directed in various ways—towards sources of sensation, towards individuals we recognize, and towards ourselves—will lead us to have a better understanding of what having a first-person perspective in all its complexity really amounts to.

Baker also provides in-depth critiques of various attempts to provide a naturalist-friendly account of the first-person perspective, and claims to show that they all fail. Her
basic strategy is to insist that: “What is needed for naturalization of the I*-concept is a third-person characterization of the capacity to conceive of oneself as oneself* in the first-person. I have argued that we have no such third-person characterization” (Baker 2013, 122).

I have a different take on the matter. I think that what this criticism demonstrates is that philosophical naturalists are in the same position as compatibilists have been with respect to indeterminists regarding free-will. Compatibilists need to reject an indeterminist reading of “could have done otherwise” and propose an alternative, determinist-friendly interpretation of this concept. So, too, some modification of what is essential for first-person perspective, other than the ontologically rich requirements that Baker insists on is probably in order. Just what those might be remains to be determined. We should not expect this to happen soon. After all, think how long it took for compatibilism to become the dominant philosophical view. We should not anticipate naturalist-friendly alternatives to Cartesian and Baker’s neo-Cartesian approaches to persons becoming popular overnight.

Natural Persons and the Problem of Personal Identity

So what does the proposed view have to say about the problem of personal identity? Simply this: as long as the same organism, O, exists over time, and as long as that organism has both the same consciousness and one of the three forms of intentionality discussed above, the same person exists. I suspect that if an organism changes which form of intentionality it has, then it may also change its identity. Again more empirical evidence about such cases is needed.

Baker would object that the naturalized account cannot handle cases in which persons switch bodies, which it seems that it cannot. But, can Baker’s own view account for survival without body switching? I have suggested that Baker’s attempt to solve the personal identity problem by appealing to the brute identity of first-person perspectives is unsatisfying. Here is another reason to consider. First-person perspectives are intentional, meaning that they are directed on objects which need not exist. Therefore, it is not unreasonable to wonder whether one’s first-person perspective from one day to the next is necessarily being directed towards the same self as itself as it was the day before. The additional element required to establish continuity is an identifying mark for a particular first-person perspective. And, that is what Baker’s view explicitly denies.
Can the naturalist view of person just sketched do any better in terms of accurately identifying a particular individual in the same body over time?

Suppose we could discover distinct natural signaling patterns occurring among the relevant parts of the brains of a statistically significant sample of research subjects, patterns that we think are responsible for the relevant mental activities to be transpiring to account for personhood status, including their engaging in first-person perspectivizing. Suppose further that we then expanded our study to include thousands of humans which further corroborated our initial findings. And, suppose even further that we could discover similar signaling going on in nonhumans accompanied by nonlinguistic but ‘pensive-like’ behaviors. What would it be reasonable for us to conclude? Baker might maintain that the signaling was merely indicative of unrelated causes at work, and further insist that we couldn’t have discovered anything remarkably unique enough to help us get a real natural reduction of the first-person perspective, let alone attribute it to nonhumans. But, what might others say about such findings? I suspect that many would not be so reticent. This scenario is, of course, pure science fiction, so no serious conclusion should be taken to follow from it. But, it might indicate the sort of evidence that future researchers will use to advance our understanding of continuing persons over time.

Another Worry to Consider

There is a final concern to address: what, if anything, is so important about the term person that it is worth trying to hang on to it within the naturalist framework that has been sketched above? After all, there are now many different things that seem to qualify as persons. A quick answer is that conferring personhood still confers moral standing. But, more needs to be said. Here’s a stab at a fuller answer to this challenge. I have claimed that there are three different kinds of intentionality corresponding to three different kinds of persons. The kind of person one is makes a difference. Unlike organisms with basic intentionality, those organisms with enhanced intentionality have the ability to recognize different individuals, typically though not restricted to members of one’s own species, including that the organism in question is itself an individual. The acquisition of enhanced intentionality is the minimal requirement for something’s being what we might call a social person, an entity that can interact with others as others. Individuals with consciousness and minimal intentionality are basic persons, individuals who can

21. I am indebted to Elizabeth Schechter for raising this concern.
interact with the environment. Individuals with fully developed first-person perspectives, on the other hand, not only recognize the individuals they interact with, they also have a developed understanding of themselves as actors among other actors. It is these individuals who have the wherewithal to be what we might term moral persons. And it is with respect to individuals in this last category that one can reasonably start to address concerns about responsibility. Given the above three part division, it seems appropriate to begin to answer the challenge with “it all depends…” That is, it depends upon what sort of person one is talking about. I take it as obvious that identifying someone as a moral person does not need any special attention. Moral persons have duties, rights, obligations, etc. So, what about the other two cases?

Once an individual has been identified as a basic person, that is conscious, which importantly includes being sentient, and also possessing intentionality, that suffices to establish a definite moral standing for such an individual. That is, practices involving and treatments of such individuals by others possessing more advanced personhood status need to come under serious moral scrutiny. In general, inflicting pain and interfering with the intentional actions of such individuals requires moral investigation and justification by moral persons. On the other hand, social persons, who have moral standing as well, also are entitled to engage in the social practices and to receive the social considerations that exist within the particular group of which they are a member.

**Final Comment**

I have devoted my efforts up to this point to outlining views I agree with partially and disagree with substantially, and to trying to provide solid criticisms of views I reject and what support I can for my own, very rough naturalistic account of persons. But, I must end by expressing my deep appreciation for the contributions of Olson, Warren, and Baker, especially, of course, those of Baker. For she has forced us to reconsider a number of claims that were taken for granted: identity vs. constitution, what’s really essential for personhood, who really has it, is it metaphysically special, and if not, how can we account for the first-person perspective. While I suspect that a number of her views regarding persons may not survive long into the future as viable options—the arc of naturalism seems to resemble the famous arc of justice—we will not have done an
adequate job of finally naturalizing persons until we can convincingly deal with the many issues that she raised.  

References


22. For further discussion of many aspects of Baker’s philosophical work, see Oliveira and Corcoran (2021).


Brutal Personal Identity

Peihong (Karl) Xie
SUNY University at Buffalo

Biography
Peihong (Karl) Xie is presently a PhD student in the Department of Philosophy, SUNY at Buffalo. His current interest is in metaphysics (especially of time and personal identity), metaphilosophy, applied ontology, and philosophy of science. He will recently publish Chinese translations of two philosophy books, which are *Metaphysics: An Introduction* by Alyssa Ney (Routledge, 2014), and *Essays and Reviews: 1959-2002* by Bernard Williams (Princeton University Press, 2014, with other two translators).

Acknowledgments
I have benefited from very detailed comments on early drafts of this paper from David Hershenov, Maureen Donnelly, Lewis Powell, Joshua Vonderhaar, and Finn Wilson. The paper was presented in the Persons Conference in Flint, Michigan, where Aderemi Artis, Simon Cushing, Mark Greene, Eric Kraemar, Elizabeth Schechter, and Jaron Cheung provided very insightful criticisms and suggestions.

Publication Details

Citation
Abstract
This paper presents a new anti-criterialist view, Brutal Personal Identity (BPI). According to BPI, personal identity is a quasi-fundamental fact, which is metaphysically grounded in brute facts about absolutely fundamental personhood. By reversing the order of metaphysical explanation, BPI is not a form of identity mysticism as Dean Zimmerman asserts. Instead, BPI has even the potential to lay a solid foundation for developing an appropriate account of mentality and first-person perspectives. Furthermore, a comparison between BPI and soul theory is provided to show why BPI is better than its main anti-criterialist rival. This provides us a compelling reason for considering BPI in the debate over personal identity.

Keywords
Anti-criterialism, Brutal Personal Identity, Natural Kind, Ontic Naturalism, Mysterious Identity, Soul Theory, Fission

Is there a non-trivial criterion for personal persistence, or personal identity over time? Criterialists' answer is simply “Yes”: There is a non-trivial criterion for personal identity. By contrast, anti-criterialists usually answer, “No.” Generally speaking, anti-criterialism is the thesis that there is no non-trivial criterion for personal identity.

While most anti-criterialists in the literature are soul theorists, this paper aims to defend a new anti-criterialist view, Brutal Personal Identity (BPI). It argues that there is no necessarily true and non-trivial criterion for personal identity just because of brutal personhood. Such an account of personal identity has significant theoretical virtues, including ontological parsimony, alignment with ontic naturalism, and respect for important modal considerations. As a result, we should seriously consider its theoretical potential.

Accordingly, this paper is divided into six sections. Section 1 and 2 briefly look through two core questions and the debate between criterialism and anti-criterialism in the philosophical context of personal identity. Then, Section 3 gives a precise formulation of BPI, and clarifies its modal significance. To motivate BPI, Section 4 assesses and rejects
a challenge from mysterious identity. Then, Section 5 provides a comparative justification for BPI by showing its superiority over typical versions of soul theory. Finally, Section 6 concludes this paper.

1 Two Core Questions of Personal Identity

As Peter van Inwagen’s (1990) makes a famous distinction between the General Composition Question and the Special Composition Question in the metaphysical context of mereology, we can make a similar distinction between two core questions in the philosophical context of personal identity. The first core question is a conceptual question, which can be called “the General Persistence Question” (GPQ):

What is the correct analysis (or engineering) of the concept of personal identity over time?

A supposed solution to GPQ provides either an analytical definition of the concept of personal identity, or a proposal about how we should use that concept for theoretical or practical purposes. In contrast, a more substantive question, which can be called “the Special Persistence Question” (SPQ), is more directly concerned with the reality of personal identity, so to speak. It is:

If a person $x$ exists at $t$ and some entity $y$ exists at $t'$, under what circumstances is it the case that $x$ is identical with $y$?

A supposed solution to SPQ has the following standard form:

(F) Necessarily, for any person $x$ existing at $t$ and any entity $y$ existing at $t'$, $x$ is identical with $y$ iff $x$ satisfies a certain criterion $\phi$ with $y$.

(F) is symbolized as:

$$(F') \Box \forall x \forall y \left( Person(x) \rightarrow ((x=y) \leftrightarrow \phi xy) \right).$$

It is not difficult to see that the condition $\phi$ is a (metaphysically) necessary and sufficient condition for personal identity.

1. I omit the formalization of temporal parameters in (F) for the purpose of simplicity. A more precise formalization is something like: $\Box \forall x \forall y \forall t \forall t' \left( Person(x, t) \land ExistsAt(x, t) \land ExistsAt(y, t') \rightarrow (x=y \leftrightarrow \phi xy) \right)$. However, this does not make a significant difference on our following discussion. So for a similar consideration, I will also omit the formalization of temporal parameters when I formalize a sufficient or necessary condition for personal identity over time (as (B1) and (B2), see Section 5.3).
It should be noted that SPQ does not ask what it takes for a person to persist as a person. Rather, it asks what it takes for a person like you or me to persist in any way at all. This formulation is thereby able to cover the theoretic possibility of animalism—the view that a human person is identified with a human animal—and other views that we are merely contingently persons. In this paper, I will focus more on SPQ than on GPQ.

2 Criterialism versus Anti-criterialism

Given SPQ and the form of a supposed criterion for personal identity, we come to the debate between criterialism and anti-criterialism. According to a standard formulation, anti-criterialism is the denial of criterialism, which is the view that there is a criterion for personal identity that is true, non-trivial, and finite.

Here are some necessary elucidations of criterialism. First, the supposed criterion for personal identity is non-trivial in the sense that it does not presuppose notions of person or personal identity in a question-begging way. Second, a statement of the supposed criterion is finitely long, so an enumeration of infinitely many individual cases of personal identity would be an inappropriate solution to SPQ in the view of criterialists.

In the current literature, most (but not all) forms of the complex view—the view that personal identity consists in some sort of qualitative continuity—are classified within criterialism. For example, most believers of the physical/psychological continuity view are criterialists. Of course, animalists are also criterialists, given my formulation of SPQ.

On the other hand, nearly all forms of the simple view—the view that personal identity is a further fact beyond any qualitative continuity—are classified within anti-criterialism. This includes soul theory, my BPI account (see below), and Bernard Williams’ (1973) view that personal identity is a further fact but bodily continuity is necessary for it.

Here a tricky case is how to classify Derek Parfit’s (1984) complex view. In his view, personal identity consists in psychological continuity, but it is not what matters at least in some cases (Parfit 1984, 217).

His famous claim that personal identity sometimes does not matter has two senses. In the ethical sense, personal identity does not always matter because it is not always what grounds one’s rational egoistic concern about one’s future. Put another way, personal identity is not – to use Jeff McMahan’s locution – a “prudential unity relation” (McMahan 2002, 42). However, the ethical sense of Parfit’s claim is based on its metaphysical sense. In the metaphysical sense, personal identity does not always
matter because SPQ is sometimes an empty question in cases like fission. In those cases, different solutions to SPQ are nothing but different descriptions of the same set of facts there, so they have no factual difference. This is why Parfit says, “we should not try to decide between the different criteria of personal identity” (Parfit 1984, 241).

In my view, Parfit’s complex view is also a special form of anti-criterialism. Some might disagree with my classification because what really makes SPQ empty in Parfit’s sense is semantic indecision. That is, it is semantically indeterminate which sort of entities in our ontology is the official referent of the word “person”, but this does not preclude those candidate sorts of entities from each having a non-trivial criterion for its diachronic identity.2

Suppose for the sake of argument that my opponents are right about the semantic indecision. Then, what kind of solution to SPQ can Parfit give? Given the semantic indecision, a possible solution is supposed to be disjunctive at best: “The correct criterion of personal identity over time is ($C_1$ or $C_2$ or $C_3$ or…)

However, such a disjunctive solution is very probably infinitely long, thus violating criterialists’ requirement of finiteness. In light of this, it is safe to classify Parfit’s (1984) view as a form of anti-criterialism. This implies that the distinction between criterialism and anti-criterialism might not coincide with the distinction between the complex view and the simple view.

3 Brutal Personal Identity

3.1 Fundamental Personhood

Now it is time to visit my BPI account. BPI is made up of four distinct theses: Person Fundamentality, No Further Explanation of Personal Identity, Necessary-Condition Contingency, and Sufficient-Condition Contingency.

The first thesis of BPI is as follows:

**Person Fundamentality:** The kind Person is an absolutely fundamental natural kind, and its kind membership is primitive. So whether an entity existing at $t$ is a person, a member of the kind Person, is a brute fact.

---

2. I’m thankful to David Hershenov for bringing up this point in personal correspondence.
Here are some necessary elucidations. Generally speaking, natural kinds are kinds that carves the nature at its joints, constituting an objective and theory-independent partition of reality. However, a natural kind can be either absolutely fundamental or not. A natural kind is absolutely fundamental if the fact that an entity is a member of it is not metaphysically grounded in any other fact. Otherwise, it is a non-fundamental natural kind. Only absolutely fundamental natural kinds are indispensable to a complete description of the whole reality. So the thesis of Person Fundamentality implies that the kind Person is indispensable to a complete description of reality.

Although the kind Person is absolutely fundamental in the above sense, this does not mean that an individual person is an absolutely fundamental entity that is similar to an individual top quark (if the standard model of physics is correct). Rather, an individual person can be something like you or me, instantiating lots of physical and/or mental properties. But neither the physical nor the mental metaphysically grounds its personhood. Whether an individual entity instantiates personhood, or the kind property of being a person, is a further fact that is as fundamental as, or even more fundamental than, physical or mental facts. Put another way, fixing its instantiation of all qualitative properties except personhood, an individual entity may be a person, or may not be a person.

Since it is brute (given Personal Fundamentality) whether an entity existing at $t$ is a person, and facts about personal identity over time have to involve the instantiation of brutal personhood, it is natural (though not logically deductive) to assert the second thesis of BPI:

**No Further Explanation of Personal Identity**: For any entity $x$ existing at $t$ and any entity $y$ existing at $t'$, if $x$ is the same person as $y$, then there is no further non-trivial explanation of the fact that $x$ is the same person as $y$ except brutal personhood.

### 3.2 Two Contingency Theses

The third and fourth theses of BPI are two contingency theses as follows:

**Necessary-Condition Contingency**: A non-logically true necessary condition for personal identity, if any, only contingently holds. In formalism, for any non-logically true condition $\phi$,

$$\Box\forall x \forall y \left(Person(x) \land (x=y) \rightarrow \phi_{xy}\right) \rightarrow \sim \Box\Box \forall x \forall y \left(Person(x) \land (x=y) \rightarrow \phi_{xy}\right).$$
**Sufficient-condition Contingency**: A sufficient condition for personal identity, if any, only contingently holds. In formalism, for any condition $\phi$,

$$\Box \forall x \forall y (\text{Person}(x) \land \phi_{xy} \rightarrow (x=y)) \rightarrow \neg \Box \Box \forall x \forall y (\text{Person}(x) \land \phi_{xy} \rightarrow (x=y)).$$

To make sense of these two contingency theses, we have to revisit the debate between criterialism and anti-criterialism. As stated in Section 2, it is a debate about whether there is a true, non-trivial, and finite criterion for personal identity. However, it is not enough for criterialists to merely assert the existence of such a criterion. Instead, they are supposed to assert that it necessarily holds by adding a second necessity operator to the front of its symbolization. That is, the following thesis is true for criterialism:

**Criterion Necessity**: There is a criterion for personal identity that necessarily holds. In formalism, for some condition $\phi$,

$$\Box \Box \forall x \forall y (\text{Person}(x) \rightarrow ((x=y) \leftrightarrow \phi_{xy})).$$

Since anti-criterialism is the denial of criterialism, one can have two ways to be an anti-criterialist now. Either one can deny the existence of any true, non-trivial, and finite criterion for personal identity, as traditional anti-criterialists did. Or one can even accept such a criterion, but argues that it is merely contingently true. It is not difficult to see that BPIers go the second way when they are committed to Necessary-Condition Contingency and Sufficient-Condition Contingency.

However, some may argue against the two contingency theses because the modal axiom 4 ($\Box \phi \rightarrow \Box \Box \phi$) falsifies them by guaranteeing that a necessary/sufficient condition necessarily holds. But it is worth noting that BPI requires a weaker modal logic than $S4$ and thus denies the modal axiom 4. Considering that our concern is metaphysical necessity here, it is not an inappropriate move for BPIers to deny the modal axiom 4. As David Braun (2022) points out, we have reason to believe that the correct logic for metaphysical necessity is a system weaker than $S4$, otherwise our logical treatment of problem cases like the *Ship of Theseus* (another puzzle of persistence!) would lead to counterintuitive results (Braun 2022, 192-193).

### 4 A Challenge from Mysterious Identity

To motivate BPI, now let us evaluate a challenge from mysterious identity, which is much inspired by Dean Zimmerman (1998). That is, since BPI denies any qualitative continuity as the necessarily true and non-trivial criterion for personal identity, it allows for a possibility in which the person $x$ is not *numerically identical* with $y$ even if $x$ is
continuous with, or even qualitatively identical with, y in all qualitative aspects. Then, doesn’t personal identity look too mysterious according to BPI?

Indeed, BPI admits of the extreme possibility of all-encompassing qualitative continuity without personal identity. However, I will argue that this is a feature, but not a drawback, of BPI. In this section, I will divide the challenge from mysterious identity into two aspects, one metaphysical and one epistemological, and then reject them.

4.1 No Metaphysical Mystery

To say that X is metaphysically mysterious in a theory is to say that X requires, but lacks, a metaphysical explanation (or a metaphysical ground) in that theory. Here we have two cases, depending on whether X is absolutely fundamental or not:

1. If X is absolutely fundamental, then it cannot be metaphysically mysterious because it does not require any further metaphysical explanation.

2. If X is non-fundamental, it does require a metaphysical explanation. But it would not be metaphysically mysterious if it is metaphysically explained by, or metaphysically grounded in, something absolutely fundamental.

In the view of BPlers, personhood is absolutely fundamental, so it is not metaphysically mysterious. Furthermore, facts about personal identity are not metaphysically mysterious, either. It is because, although those facts are not absolutely fundamental, they are quasi-fundamental in the sense that they are directly metaphysically explained by facts about the instantiation of brutal personhood.

Given brutal personhood, a better explanation of mentality is even available to BPlers. Recall Parfit’s (1984) distinction between genuine memory and quasi-memory. A core feature of our genuine memory is that “we can remember only our own experiences.” (Parfit 1984, 202). It is clear that such a notion of memory presupposes the notion of personal identity, so an account of personal identity in terms of genuine memory is circular or question-begging. To fix the issue of circularity or triviality, Parfit (1984) invents a technical notion of quasi-memory in developing his psychological continuity view. Roughly speaking, one has quasi-memory when one seems to remember having an experience that might be someone else’s (Parfit 1984, 219-223). Similar distinctions also apply to other kinds of mental states.
However, the technical notion of quasi-memory or, more generally, quasi-mentality, seems too *ad hoc*. Luckily, BPIers can reverse the order of metaphysical explanation and thus avoid those *ad hoc* notions. That is, BPI does not require memory or other mental states to metaphysically explain personhood and thus personal identity. Rather, it is brutal personhood that (at least partially) metaphysically explains personal identity and then genuine mentality! This is why the Parfitian inventions about quasi-mentality are not necessary for BPIers.

Furthermore, it is worth noting that BPI is still neutral on how mentality works, given the above reversal of explanation order. So BPI may be compatible with any plausible philosophical account of mental mechanisms. Therefore, given BPI’s rich explanatory power and its theoretic neutrality, it is less metaphysically mysterious than my opponents suppose it is.

### 4.2 No Epistemic Mystery?

Still, someone may charge that BPI is epistemically mysterious. She may say, if BPI is true, there is a possible case that Sam fails to be the same person as Sam* even if “all of the non-branching psychological, phenomenal, physical, biological, etc., connections obtain between them” (Duncan 2020, 174). This possibility undermines our everyday knowledge about our persistence. BPI is false because “we do know that we persist!” (Duncan 2020, 177)

In fact, Matt Duncan (2020) intends to use this epistemic objection to reject all forms of anti-criterialism. Since BPI is a sort of anti-criterialism, can its proponents make any progress in resolving the above epistemological challenge?

I think BPIers can. To see how to do it, it is worth noting that Duncan’s epistemic objection could be reformulated in terms of the *relevant alternatives theory* (Rysiew 2006), the view that an epistemic agent $E$ knows that $P$ only if $E$’s total evidence is sufficient to preclude all relevant alternatives to the state of affairs which $P$ is true of.

Here are two further points about the notion of relevant alternative. First, a state of affairs $Q$ is an *alternative* to another state of affairs $K$ if $Q$ is incompatible with $K$. Second, although controversial, the alternative $Q$ is *relevant* in the general sense that $Q$ shares similar basic features with $K$ in an epistemic evaluation. For example, $Q$ is very similar to $K$ in respect of their external environment, underlying metaphysical setting, $E$’s cognitive abilities, and so on.

Now Duncan’s objection can be formulated as a skeptical argument:
(i) BPI is true. (a presumption for reductio)

(ii) If BPI is true, then it is possible that we fail to persist despite having all qualitative features that a normal persisting person has. (a corollary of premise (i))

(iii) So, it is possible that we fail to persist despite having all qualitative features of a normal persisting person. (by (i) and (ii) and modus ponens)

(iv) We know that we persist only if our total evidence is sufficient to preclude all relevant alternatives to our persistence. (the relative alternatives theory)

(v) The possibility of our failing to persist despite having all qualitative features is a relevant alternative to our persisting. (Duncan’s claim)

(vi) But our total evidence is insufficient to preclude the above possibility. (Duncan’s claim)

(vii) So, we do not know that we persist. (by (iii)-(vi) and modus tollens)

(viii) But we do know that we persist. (common sense)

Therefore,

(ix) BPI is false. (by (vii) and (viii) and reductio ad absurdum)

For BPIers, a promising approach to addressing the above argument is to deny its premise (v).

The basic idea is that the possibility of our failing to persist despite having all kinds of qualitative continuities, even if it is a genuine possibility, is still an irrelevant alternative to our persistence. Here BPIers can follow Trenton Merricks (1998, 107-109) to distinguish criterion from evidence. BPIers deny any biological, physical or psychological continuity as the necessarily true criterion for personal identity, but allow them as good though fallible evidence for personal identity. And the scope of evidence can even include the sameness
of fingerprints or clothing! So at the level of evidence, any qualitative continuity is not much deeper than the sameness of fingerprints or clothing.

In light of this, we have rich evidence for our everyday belief in our persistence. If our belief happens to be true due to worldly arrangements, then it is not unreasonable to claim that we possess everyday knowledge of our persistence in that case. The extreme possibility referred to by the premise (v) does not undermine our knowledge about our persistence in most ordinary cases just because it is a matter of extremely bad epistemic luck. Any reasonable account of knowledge is supposed to make room for such luck. Otherwise, we would have to return to the very implausible requirement of infallible knowledge as proposed by Descartes.

Nonetheless, opponents of BPI may still feel dissatisfied with the above reply because it violates the well-known KK principle that for any proposition $p$, if one knows that $p$, then one knows that one knows it (Hemp 2023). They may argue that given the BPI-based reply, even if we know that we persist, we do not know that we know that we persist. It may be because our knowledge about our persistence depends on whether it is true that we persist, and the latter is largely dependent on worldly arrangements, which are beyond our internal grasp.

In response to this, I will point out that the same charge is also against externalists of knowledge, who claim that an epistemic agent’s lack of internal access to the basis for her knowledge does not necessarily undermine her knowledge. For example, a reliabilist may argue that knowledge is true belief generated by a reliable process of some sort, but she does not require any internal access to any reliable belief-generating process to ensure knowledge. So reliabilism, as a form of externalism, also violates the KK principle. Here BPIers can take sides with externalists. Some plausible externalist conception of knowledge may be essential to BPI.

5 A Comparative Justification for BPI

In this section, I will present a comparative justification for BPI by demonstrating its superiority over its main anti-criterialist rival, soul theory. For those who are inclined towards anti-criterialism, this comparative justification would offer them a compelling reason to seriously consider BPI. Let us start with a brief elucidation of soul theory.
5.1 What is Soul Theory?

Soul theorists are usually anti-criterialists. According to soul theory, a person has essentially a soul. She may be either a composite of a body and a soul, or just a soul. Either way, however, diachronic soul identity is indispensable to personal identity. There is no non-trivial criterion for personal identity simply because there is no non-trivial criterion for soul identity.

Regarding the origin of soul, soul theorists can have two competing conceptions. On the one hand, soul naturalists take a soul as a natural but immaterial simple: Either it emerges from an alive brain of considerable complexity (Hasker 2001; Zimmerman 2010), or it has an intrinsic disposition of pairing with a certain brain to support consciousness (Unger 2006). On the other hand, traditional soul theists argue that a soul is “an individual substance of a rational nature”, which is created and implanted into a body by God (Shoemaker 2005, 56).

Many soul theorists believe that the mental nature of a soul implies that some sort of psychological continuity or (at least) psychological capacity is necessary for personal identity. Such a necessary condition for personal identity is thereby grounded in the essence of soulhood or personhood, making the following thesis true for soul theorists:

**Necessary-condition Necessity**: There is a non-logically true necessary condition for personal identity that necessarily holds. In formalism, for some non-logically true condition $\phi$, $\square\forall x \forall y (\text{Person}(x) \land (x = y) \rightarrow \phi_{xy})$.

Next, I will present how soul theorists holding Necessary-condition Necessity are confronted with four problem cases: qualitative continuum, modal coincidence, graduality, and fission. Then, I will outline how BPIers can effectively address those challenges.

5.2 Qualitative Continuum

Matt Duncan (2020) argues that all anti-criterialists should accept the presence of some non-trivial necessary conditions for personal identity. For example, I cannot persist until tomorrow if the universe will be destroyed before then. So it is a necessary condition for my persistence until tomorrow that the universe will not be destroyed before then (Duncan 2020, 6).
Some ambitious soul theorists are not content with accepting such necessary conditions. Rather, they believe there is at least some necessary condition for personal identity that is essence-grounded and thus necessarily true. In particular, many (but not all) of them think that Necessary-condition Necessity is at least true of some sort of psychological continuity. An objection from psychological continuum, however, would show why it is not the case.

A psychological continuum is a range of possible cases covering all possible degrees of some sort of psychological continuity, which could be an overlapping chain of a certain amount of memories, desires, or other psychological states. Now at the near end of the psychological continuum in question, there is a person called Sam. Step by step, Sam’s psychological continuity will reduced to a lesser and lesser degree, so that in the far-end case, there would be another entity Sam* that is not psychologically continuous with Sam at all, though continuous with Sam in all other qualitative aspects. Such diachronic changes concerning Sam’s psychological continuity are presented in a series of intermediary cases connecting the near-end case and the far-end case. The entities in any two adjacent cases are duplicates of each other except there is an extremely slight difference between their psychologies. So it seems natural to say that if the entity in one case is Sam, then the extremely slightly different entity in another adjacent case would also be Sam.

But if so, a simple proof by mathematical induction will show that Sam* is Sam:

(1) Inductive base: The person in the near-end case is Sam.

(2) Inductive step: If the entity in one case is Sam, then the entity in another adjacent case is also Sam.

Therefore, by mathematical induction,

(3) The entity Sam* in the far-end case is Sam.

However, Sam* is not psychologically continuous with Sam at all. Therefore, the above proof implies that given the possibility of the psychological continuum, the psychological continuity in question is not necessary for personal identity. So it is not necessarily the case that the psychological continuity in question is necessary for personal identity, showing that Necessary-condition Necessity is false of the psychological continuity in question. Similar reasonings can be easily extended to any other sort of qualitative continuity.
Of course, the above reasoning can be rejected by denying the inductive step (2). That is, it is possible that there is an “abrupt change” happening in some two adjacent cases so that the entity is Sam but the entity in another adjacent case is not. But it is difficult to see how the “abrupt change” is a necessity. So if it is possible that the “abrupt change” does not happen within a psychological continuum, the above reasoning against Necessary-condition Necessity remains valid, posing a challenge for soul theorists.

Nevertheless, BPI is immune from the objection from qualitative continuum because it is merely committed to Necessary-Condition Contingency. For BPIers, any sort of qualitative continuity, even if necessary for personal identity, is only contingently necessary for it. For example, the sort of psychological continuity required by a soul theorist may be necessary for personal identity in some cases, but not in other cases like the psychological continuum in question. So BPIers do give a solution to the problem of qualitative continuum, which is much more elegant than other solutions available to soul theorists holding Necessary-condition Necessity.

5.3 Modal Coincidence

Duncan (2020) claims that there are three key motivations against criterialism: Merricks’ (1998) argument from modal coincidence, the argument from graduality, and the argument from fission. He also points out, however, that any anti-criterialist would be subject to the same charges if she accepts the existence of a non-trivial sufficient condition for personal identity. Therefore, he concludes that all anti-criterialists should deny any non-trivial sufficient condition.

Take Merricks’ argument from modal coincidence first. As Section 1 shows, a standard solution to SPQ is supposed to have the following form:

\[(F') \Box \forall x \forall y (\text{Person } (x) \rightarrow (x = y \leftrightarrow \phi xy)).\]

However, Merricks (1998) argues that a standard solution of this form in fact requires criterialists to establish a necessary connection between two contingent states of affairs: one state of affairs is the person x at t’s being identical with the entity y at t’, and another is x’s satisfying the supposed criterion \( \phi \) with y. However, such a necessary connection between two contingent states of affairs does not look very intuitive. Why should we believe in the first place that there is any necessary connection between two contingent entities? Isn’t it more probable that they have only some contingent relationship (Merricks 1998, 116-118)?
Following Merricks, Duncan argues that the same argument, if appropriate, could also be used against anti-criterialists who adhere to the existence of some non-trivial sufficient condition for personal identity. For if $\phi$ is a sufficient condition satisfied by $x$ and $y$, then the following symbolization holds true:

$$\Box \forall x \forall y (\text{Person}(x) \land \phi xy \rightarrow (x = y)).$$

This seems to be another case of a necessary connection between two contingent states of affairs. So Duncan concludes that anti-criterialists should not accept any sufficient condition for personal identity (Duncan 2020, 8).

Indeed, hardly any anti-criterialist actually acknowledges a sufficient condition for personal identity. Nevertheless, Duncan’s reasoning is flawed because he does not find that his criticism of sufficient condition also applies to any necessary condition for personal identity. For if there is a certain necessary condition $\phi$ for personal identity, the following necessary connection holds between two relevant contingent states of affairs:

$$\Box \forall x \forall y (\text{Person}(x) \land (x = y) \rightarrow \phi xy).$$

So if Duncan’s conception of modal coincidence is correct, then anti-criterialists including soul theorists have to deny any necessary condition for personal identity, either. Unfortunately, this corollary is obviously inconsistent with Duncan’s earlier claim in his paper (2020) that anti-criterialists should acknowledge at least some non-trivial necessary conditions for personal identity (recall the first paragraph in Section 5.1)!

The internal inconsistency in Duncan’s claims suggests that he misses the point of Merricks’ argument. There is no problem with a necessary connection between two contingent states of affairs. Rather, it really matters whether the necessary connection in question is well-grounded.

In fact, nearly all criterialists argue that the necessary connection involved in (F$'\phi$) is well-grounded: it is grounded in the essence of personhood (or, for example, animalhood for animalists)! This is why they tend to accept the thesis of Criterion Necessity. Similarly, many soul theorists also claim that the necessary connection involved in (B2) is grounded in the essence of personhood or soulhood. So they tend to accept the thesis of Necessity-Condition Necessity.

By contrast, BPIers deny Criterion Necessity or Necessity-Condition Necessity because they do not think that any non-trivial necessary and/or sufficient condition for personal identity is essence-grounded.

It is worth noting that there is something subtle here. BPIers, just like many soul theorists, deny any essence-grounded sufficient condition for personal identity. But
unlike many soul theorists, they can still acknowledge the presence of some *contingently* sufficient condition in some cases! In those cases, they are indeed committed to a necessary connection between two contingent states of affairs. But since Sufficient-condition Contingency shows that such a necessary connection only contingently obtains, BPlers would not be thereby in a worse situation than criterialists or soul theorists.

5.4 Graduality

Another motivation against criterialism is said to come from *graduality*. The idea is very intuitive: Personal identity is *all-or-nothing*, whereas many candidate conditions serving as criteria for personal identity *admit of degrees*. Therefore, to establish a criterion for personal identity, criterialists have to determine a precise threshold above which the supposed condition is met for a person to persist. For example, Parfit (1984) defends his psychological criterion by requiring an overlapping chain of “*strong connectedness*”, which involves *at least half* of the psychological connections between any two times at which a normal person has (Parfit 1984, 206). However, such determination of a threshold is undoubtedly arbitrary.

Duncan (2020) argues that an anti-criterialist would be subject to the same objection if she acknowledges some non-trivial sufficient condition for personal identity. In that case, she has to determine an arbitrary threshold for a certain condition admitting of degrees to be sufficient for personal identity. However, if anti-criterialists should deny any non-trivial sufficient condition for this reason, why shouldn’t they deny any non-trivial necessary condition for a similar consideration? After all, needn’t they also determine a threshold for a certain condition admitting of degrees to be necessary for personal identity? So Duncan’s claim is again inconsistent with his earlier claim that anti-criterialists should acknowledge at least some non-trivial necessary condition for personal identity.

Here two contingency theses involved in BPI are conducive to addressing the above graduality problem. While a criterialist has trouble in determining a precise threshold for the *unique* criterion for personal identity, a BPler is free to acknowledge that there are different sufficient/necessary conditions in different cases, each of which has a certain threshold. There is no need for a further explanation why a sufficient/necessary condition has the threshold it has in a certain case. It is simply a contingent brute fact in reality.
5.5 Fission

5.5.1 A Problem for Criterialism

Finally, we reach the most important motivation against criterialism: fission. Suppose a person, say, Bruce, undergoes fission, resulting in two distinct persons, Lefty and Righty, who are two nearly perfectly qualitative duplicates. The same striking amount of qualitative connections obtain not only between Bruce and Lefty, but also between Bruce and Righty, so Lefty and Righty seem equally good candidates for being Bruce. If the amount of qualitative connections in question is the criterion for personal identity, then Bruce would be not only identical with Lefty, but also identical with Righty. And then we can infer from the symmetry and transitivity of identity that Lefty is identical with Righty. But it is clear that they are two distinct persons – a contradiction.

It is worth noting that it won’t help criterialists very much if they argue that the qualitative continuity in question constitutes a criterion for personal identity only when it is non-branching. It is because the non-branching constraint makes the personal identity of Bruce and, say, Lefty, dependent on an extrinsic matter of whether a third candidate, say, Righty, is present. However, it is more reasonable to argue that the personal identity of $x$ and $y$ only depends on their internal relationship. In light of this, many criterialists have attempted to directly refine their criterion, finally leading to three theoretic options available to them: either that Bruce would cease to exist after fission, or he would be identical with either Lefty or Righty, but not both.

5.5.2 Two Solutions from Soul Theory

Duncan (2020) points out that the same contradiction would be generated again even if the amount of qualitative connections in question is not a criterion but merely a sufficient condition for personal identity. Therefore, he argues that anti-criterialists have to deny any non-trivial sufficient condition for personal identity (Duncan 2020, 8-9).

In fact, this is exactly what soul theorists usually do in the case of fission. In their view, no qualitative continuity is sufficient for personal identity. So it is not the case that Bruce would be identical with two different post-fission persons. Rather, Bruce would be at most identical with only one of the post-fission persons.

Following the above line of argument, two distinct solutions are available to soul theorists. A soul naturalist might say, the soul inhabited in Bruce’s body goes with one of the new bodies, say, Lefty’s body, while a new soul emerges from, or pairs with, Righty’s body. Since Lefty and Righty are nearly perfectly qualitative duplicates, however,
why isn’t it the opposite case that the original soul goes with Righty’s body and a new soul emerges from, or pairs with, Lefty’s body? So a soul naturalist has to rely on some naturalistic process to prevent this alternative from occurring, but often such a naturalistic account is lacking.

“No further explanation is needed.” A soul theist says so, on the other hand. She may argue that it is God that chooses the original soul to go with Lefty’s body while creating a new soul inhabited in Righty’s body. There is no further explanation of God’s choice because His choice is “like us considering which of two qualitatively identical snacks to eat” (Hershenov and Taylor 2014, 25, endnote 11).

5.5.3 A BPI-based Solution

Then, what can BPIers say about fission? Unlike many soul theorists, BPIers needn’t deny all sorts of sufficient conditions for personal identity. Rather, they may accept some contingently sufficient condition, and then claim that the amount of qualitative connections in question, though sufficient for personal identity in other cases, is insufficient in Bruce’s case. So Bruce’s case is not that he is identical with two different post-fission persons. Even if Lefty and Righty look like equally good candidates for being Bruce, at most one of them, say, Lefty, is in fact identical with Bruce. When asked why it is Lefty but not Righty that is identical with Bruce, BPIers could happily answer, “No further explanation. It is just a fact grounded in brutal personhood.”

This solution based on BPI falls between the solution proposed by soul naturalism and the solution presented by soul theism. BPI is a form of ontic naturalism because it asserts that the kind Person is a natural kind. However, unlike soul naturalists, BPIers do not owe us a further naturalistic explanation about “why not the opposite”. In their view, the question of whether Bruce’s soul goes with Lefty or Righty does not require any further explanation beyond brutal personhood. In this sense, BPI is better than soul naturalism when facing fission.

On the other hand, it is clear that the BPI-based solution is more similar to the solution proposed by soul theists because they both deny any further explanation of personal identity in Bruce’s case. However, they have a substantial difference in their ontological posits. While BPIers posit Person as an absolutely fundamental natural kind, soul theists (in Hershenov and Taylor’s sense) posit God to prevent a further explanation. But why do we bother positing God if a fundamental-kind posit of Person has been enough to provide at least an equally good elucidation of personal identity? Here a methodological principle is: if two distinct ontological posits are equally good
in explaining the same set of phenomena, *ontological parsimony* always requires us to choose the sparser one rather than the richer one. So according to this principle, BPI does a better job in explaining fission than soul theism.

Based on the above considerations, I conclude that BPI is better than common versions of soul theory when facing fission.

6 Conclusion

In this paper, I give a precise formulation of BPI, the view that personal identity is a quasi-fundamental fact, which is metaphysically grounded in facts about absolutely fundamental personhood. Such an account of personal identity is immune from the metaphysical challenge from mysterious identity, and it can even get rid of a general epistemic objection to anti-criterialism if it buys some plausible externalist conception of knowledge.

Moreover, a comparative justification is available to BPIers by demonstrating that BPI is better to address four problem cases than common versions of soul theory. Of course, such a justification is not decisive. But at least it shows that BPI is a very attractive option in the debate over personal identity, so it should not be so easily ignored.

References


