Journal of Cognition and Neuroethics

Journal of Cognition and Neuroethics

Table of Contents

Journal of Cognition and Neuroethics

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

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

Citation

Cheung, Jaron J. 2023. "Holism, Narrative, and Paradox: New Criteria for Settling Disputes in Personal Identity." *Journal of Cognition and Neuroethics* 9 (2): 1–20.

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.

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

2002, 293–294). 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 cephalothoracopagus 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 another), 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

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

Campbell, Tim, and Jeff McMahan. 2010. "Animalism and the Varieties of Conjoined Twinning." *Theoretical Medicine and Bioethics* 31 (4): 285–301.

Condic, Maureen L. 2013. "When Does Human Life Begin? The Scientific Evidence and Terminology Revisited." *University of St. Thomas Journal of Law & Public Policy* 8 (1): 44–81.

Hershenov, David B. 2013. "Who Doesn't Have a Problem of Too Many Thinkers?" *American Philosophical Quarterly* 50 (2): 203–208.

Liao, S. Matthew. 2006. "The Organism View Defended." *The Monist* 89 (3): 334–350.

MacIntyre, Alasdair C. 2007. *After Virtue: A Study in Moral Theory*. Third edition. Notre Dame, Ind: University of Notre Dame Press.

McMahan, Jeff. 2002. *The Ethics of Killing: Killing at the Margins of Life*. New York: Oxford University Press.

Ricoeur, Paul. 1992. *Oneself As Another*. Chicago: University of Chicago Press.

Schechtman, Marya. 2014. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. First edition. Oxford: Oxford University Press.

Schechtman, Marya. 1996. *The Constitution of Selves*. Ithaca, NY: Cornell University Press.

Journal of Cognition and Neuroethics

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

Jeffrey P. Fry
Ball State University

Citation
Fry, Jeffrey P. 2023. "Is Anyone on First? Sport, Agency, and the Divided Self." *Journal of Cognition and Neuroethics* 9 (2): 21–34.

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.

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" (Dennett[2] 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 Some*one* 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).[5] 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]

---

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

---

have different goals and different values, The fractionation may not be revealed by brain injury, but it also underlies everyday behavior" (Levy 2018, 118).

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).[7] 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).[9]

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

Bargh, John. 2017. *Before You Know It: The Unconscious Reasons We Do What We Do*. New York: Touchstone/Simon & Schuster, Inc.

Blackmore, Susan. 2017. *Consciousness: A Very Short Introduction*. 2nd ed. New York: Oxford University Press.

Dennett, Daniel C. 2003. "Who's On First? Heterophenomenology Explained." *Journal of Consciousness Studies*, 10, No. 9–10 19–30.

Dennett, Daniel C. 2013. *Intuition Pumps and Other Tools for Thinking*. New York: W.W. Norton & Company.

Dennett, Daniel C. 2017. *From Bacteria to Bach and Back*. New York: W.W. Norton & Company.

Descartes, René. [1641] 1993. *Meditations on First Philosophy in Which the Existence of God and the Distinction Between the Soul From the body Are Demonstrated*. 3rd ed., translated by Donald A. Cress, Indianapolis, IN: Hackett Publishing Company.

Descartes, René. [1649] 2021. *The Passions of the Soul*. (Excerpt from translation [J. Cottingham, R. Stoothoff, and D. Murdoch, eds./trans.] published in 1996). In *Philosophy of Mind: Classical and Contemporary Readings*, 2nd ed., edited by David J. Chalmers 20–22. New York: Oxford University Press.

Eagleman, David. 2011. *Incognito: The Secret Lives of the Brain*. New York: Pantheon Books.

Farah, Martha J., and Andrea S. Heberlein. 2010. "Personhood: An Illusion Rooted in Brain Function?" In *Neuroethics: An Introduction with* Readings, edited by Martha J. Farah, 321–338. Cambridge, MA: MIT Press.

Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: The MIT Press.

Francis, Bill. June 2 2016. "Who's on First Joined the Hall 60 Years Ago." https://baseballhall.org/discover/short-stops/whos-on-first.

Frankish, Keith, editor. 2017. *Illusionism: A Theory of Consciousness*. Exeter, U.K.: Imprint Academic.

Gazzaniga, Michael S. 2016. *Tales From Both Sides of the Brain: A Life in Neuroscience*. New York: Ecco.

Gazzaniga, Michael S. 2011. *Who's in Charge? Free Will and the Science of the Brain*. New York: Ecco.

Hofstadter, Douglas. 2007. *I am a Strange* Loop. New York: Basic Books.

Hume, David. 1968. *A Treatise of Human Nature*, 2 vols., vol. I. London: Dent; New York: Dutton. Everyman's Library. Reprint. (Orig. Everyman's Library edition pub. 1911.)

Lazenby, Roland. 2014. *Michael Jordan: The Life*. New York: Little, Brown and Company.

Levy, Neil. 2018. "Choices Without Choosers: Toward a Neuropsychologically Plausible Existentialism." In *Neuroexistentialsim: Meaning, Morals, and Purpose in the Age of Neuroscience*, edited by Gregg D. Caruso and Owen Flanagan, 111–125. New York: Oxford University Press.

Metzinger, Thomas. 2004. *Being No One: The Self-Model Theory of Subjectivity*. Cambridge, MA: The MIT Press.

Papineau, David. 2017. *Knowing the Score: What Sports Can Teach Us About Philosophy (And What Philosophy Can Teach Us About Sports)*. New York: Basic Books.

Ross, Lee and Richard Nisbett. 2010. *The Person and the Situation*. Selections from *The Person and the Situation: Perspectives of Social Psychology*. (Original pub. 1991.) In *Moral Psychology: Historical and Contemporary Readings*, edited by Thomas

Nadelhoffer, Eddy Nahmias, and Shaun Nichols, 187–196. Wiley-Blackwell.

Schechter, Elizabeth. 2018. *Self-Consciousness and Split Brains: The Mind's* I. New York: Oxford University Press.

Schechter, Elizabeth. April 20 2020. Aeon. "What We Can Learn About Respect and Identity From Plurals." https://aeon.co/ideas/what-we-can-learn-about-respect-and-identity-from-plurals.

Searle, John R. 2004. *Mind: A Brief Introduction*. New York: Oxford University Press.

Sellars, Wilfred. 1962. Chapter 2, "Philosophy and the Scientific Image of Man." In *Frontiers of Science and Philosophy*, edited by Robert Colodny, 35–78. Pittsburgh, PA: University of Pittsburgh Press.

Sokolowski, Robert. 2000. *Introduction to Phenomenology*. Cambridge, U.K. Cambridge University Press.

Strawson, Galen. 2009. *Selves: An Essay in Revisionary Metaphysics*. New York: Clarendon Press/Oxford University Press.

Hassin, Ran R., James S. Uleman, and John A. Bargh, editors. 2007. *The New Unconscious*. New York: Oxford University Press.

Walker, Herschel, with Gary Brozek and Charlene Maxfield. 2008. *Breaking Free: My Life With Dissociative Identity Disorder*. New York: Touchstone and Howard Books/Simon & Schuster, Inc.

Wegner, Daniel. 2002. *The Illusion of Conscious Will*. Cambridge, MA: The MIT Press.

Journal of Cognition and Neuroethics

Persons, Person Stages, Adaptive Preferences, and Historical Wrongs

Mark E. Greene
University of Delaware

Persons, Person Stages, Adaptive Preferences, and Historical Wrongs

Mark E. Greene

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.

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

treating 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 re-states 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

Barnes, Elizabeth. 2009. "Disability and Adaptive Preference." *Philosophical Perspectives* 23 (1): 1–22.

Greene, Mark, and Steven Augello. 2011. "Everworse: What's Wrong with Selecting for Disability?" *Public Affairs Quarterly* 25 (2): 131–140.

Harman, Elizabeth. 2009. "Harming as Causing Harm." In *Harming Future Persons: Ethics, Genetics and the Nonidentity Problem*, edited by Melinda A Roberts and David Wasserman, In International Library of Ethics, Law, and the New Medicine, 137–154. Dordrecht: Springer.

Human Fertilization and Embryology Act. 2008. U.K.

Hume, David. 1888. *A Treatise of Human Nature*. edited by L. A. Selby-Bigge: Clarendon Press.

Katz, Leo. 1996. *Ill-Gotten Gains: Evasion, Blackmail, Fraud, and Kindred Puzzles of the Law*. Chicago: Chicago University Press.

Lewis, David. 1976. "Survival and Identity." In *The Identities of Persons*, edited by Amélie Oksenberg Rorty, 17–40. Berkeley: University of California Press.

Moller, Dan. 2011. "Wealth, Disability, and Happiness." *Philosophy & Public Affairs* 39 (2): 177–206.

Morris, Christopher W. 1984. "Existential Limits to the Rectification of past Wrongs." *American Philosophical Quarterly (Oxford)* 21 (2): 175–182.

Nussbaum, Martha. 2003. "Capabilities as Fundamental Entitlements: Sen and Social Justice." *Feminist Economics* 9 (2-3): 33–59.

Parfit, Derek. 1982. "Future Generations: Further Problems." *Philosophy and Public Affairs* 11 (2): 113–172.

Sen, Amartya. 1985. *Commodities and Capabilities. Professor Dr. P. Hennipman Lectures in Economics v. 7*. Amsterdam, North-Holland.

Shiffrin, Seana Valentine. 2012. "Harm and its Moral Significance." *Legal Theory* 18 (Special Issue 03): 357–398.

Thomas, Roy. 1966. "I, Lucifer!" *The X-Men*, May 20.

Journal of Cognition and Neuroethics

Naturalizing Persons: A Reply to Lynne Rudder Baker†

Eric Kraemer
University of Wisconsin-La Crosse

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.

Naturalizing Persons: A Reply to Lynne Rudder Baker

Eric Kraemer

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.[1] 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.[5] 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).[6]

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.[7] 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.[9] 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.[10] 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 1st-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.[13] 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.[14]

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).[15] 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?[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

---

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?

competent humans and also of granting there are still a number of unclear issues that need to be acknowledged.

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)[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.[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.[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.'[20] 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[21]

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.[22]

References

Baker, Lynne Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge: Cambridge University Press.

Baker, Lynne Rudder. 2005. "Death and the Afterlife." In William Wainwright, editor, *The Oxford Handbook of Philosophy of Religion*, 365–391. Oxford: Oxford University Press.

Baker, Lynne Rudder. 2007. *The Metaphysics of Everyday Life*. Cambridge: Cambridge University Press.

Baker, Lynne Rudder. 2013. *Naturalism and the First-Person Perspective*. Oxford: Oxford University Press.

Baker, Lynne Rudder. 2016. "Making Sense of Ourselves: self-narratives and personal identity." *Phenomenology and the Cognitive Sciences* 15 (1):7–15.

Chisholm, Roderick M. 1956. "Sentences about Believing." *Proceedings of the Aristotelian Society, New Series* 56: 125–148.

Davidson, Donald. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association* 60 (3): 441–458.

Godfrey-Smith, Peter. 2017. *Other Minds: The Octopus, the Sea, and the Deep Origin of Consciousness*. New York: HarperCollins.

Kraemer, Eric Russert. 1983. "Abortion and Cloning." *The Southern Journal of Philosophy* 21 (4):537–545.

Kraemer, Eric Russert. 2006. "Moral Mysterianism." *Southwest Philosophical Review* 22 (1): 69-77.

Noonan, John. 1968. "Deciding Who Is Human." *Natural Law Forum* 13 (1):134–140.

Oliveira, Luis and Kevin Corcoran. 2021. *Common Sense Metaphysics—Essays in Honor of Lynne Rudder Baker*. New York: Routledge.

Olson, Eric T. 1997. *The Human Animal*, New York: Oxford University Press.

Parfit, Derek. 1986. *Reasons and Persons*. New York: Oxford University Press.

---

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

Strawson, P. F. 1958. "Persons." *Minnesota Studies in the Philosophy of Science, Volume II*, edited by Herbert Feigl, Michael Scriven, and Grover Maxwell, 330–354. Minneapolis: University of Minnesota Press

Unger, Peter. 1979. "Why there are no people." *Midwest Studies in Philosophy* 4 (1): 177–222.

Warren, Mary Anne. 1973. "On the Moral and Legal Status of Abortion." *The Monist* 57 (1): 43–69.

Warren, Mary Anne. 1984. "Postscript." In *The Problem of Abortion*, edited by Joel Feinberg, 2nd Edition, 116-119. Belmont: Wadsworth.

Journal of Cognition and Neuroethics

Brutal Personal Identity

Peihong (Karl) Xie [ORCID]
SUNY University at Buffalo

Brutal Personal Identity

Peihong (Karl) Xie

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 ϕ with y.

(F) is symbolized as:

> (F′) $\Box\forall x\forall y$ (*Person* $(x)\rightarrow((x=y)\leftrightarrow\phi xy))$.[1]

It is not difficult to see that the condition ϕ 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: $\Box\forall x\forall y\forall t\forall t'$(*Person* $(x, t)\wedge ExistsAt(x, t)\wedge ExistsAt(y,t')\rightarrow(x=y\leftrightarrow\phi xy))$. 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…)", where C_n stands for the persistence criterion of some sort of entities, which is as a candidate referent of the word "person". 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 ϕ, $\Box \forall x \forall y \, (Person \, (x) \land (x=y) \rightarrow \phi xy) \rightarrow \sim \Box\Box \forall x \forall y \, (Person \, (x) \land (x=y) \rightarrow \phi xy)$.

> **Sufficient-condition Contingency**: A *sufficient* condition for personal
> identity, if any, only contingently holds. In formalism, for any condition φ,
> □∀x∀y (*Person* (x)∧φxy→(x=y)) → ~ □□∀x∀y (*Person* (x)∧φxy→(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 φ,
> □□∀x∀y (*Person* (x)→((x=y)↔φxy)).

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** (□φ → □□φ) 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 BPIers, 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 BPIers. 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 ϕ, $\Box\Box\forall x\forall y$ (*Person* $(x)\wedge(x=y)\rightarrow\phi xy$).

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

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 ϕ 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 φ is a sufficient condition satisfied by x and y, then the following symbolization holds true:

(B1) $\Box\forall x\forall y$ (*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 φ for personal identity, the following necessary connection holds between two relevant contingent states of affairs:

(B2) $\Box\forall x\forall y$ (*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′) 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, BPIers 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 BPIer 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.

<u>5.5 Fission</u>

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

Braun, David. 2022. *Logical Methods in Philosophy*. Manuscript.

Duncan, Matt. 2020. "A Renewed Challenge to Anti-criterialism." *Erkenntnis* 85 (1). URL = ‹https://doi.org/10.1007/s10670-018-0023-7›.

Hasker, William. 2001. *The Emergent Self*. London: Cornell University Press.

Hemp, David. "The KK (Knowing that One Knows) Principle." *The Internet Encyclopedia of Philosophy*, ISSN 2161-0002, URL=‹ https://iep.utm.edu/kk-princ/ ›, June 22, 2023.

Hershenov, David, and Adam Taylor. 2014. "Split Brains: No Headache for the Soul Theorist." *Religious Studies*, 50(4). URL = ‹https://web.archive.org/web/20170330221306id_/http://www.acsu.buffalo.edu/~dh25/articles/SplitBrainsNoHeadacheForSoulTheorists.pdf›.

McMahan, Jeff. 2002. *The Ethics of Killing: Problems at the Margins of Life*. New York: Oxford University Press.

Merricks, Trenton. 1998. "There Are No Criteria of Identity over Time." *Noûs* 32 (1): 106–124.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Rysiew, Patrick. 2006. "Motivating the Relevant Alternatives Approach." *Canadian Journal of Philosophy* 36: 259–279.

Shoemaker, David W. 2005. "Embryos, Souls, and the Fourth Dimension." *Social Theory and Practice* 31(1): 51–75.

Unger, Peter. 2006. *All the Power in the World*. New York: Oxford University Press.

van Inwagen, Peter. 1990. *Material Beings*. Ithaca, N.Y.: Cornell University Press.

Williams, Bernard. 1973. *Problems of the Self*. Cambridge: Cambridge University Press.

Zimmerman, Dean. 1998. "Criteria of Identity and the 'Identity Mystics'". *Erkenntnis* 48(2–3): 281–301.

Zimmerman, Dean. 2010. "From Property Dualism to Substance Dualism." *The Proceedings of the Aristotelian Society* LXXXIV: 119–150.

Introducing Plurals

Elizabeth Schechter
University of Maryland, College Park

Introducing Plurals

Elizabeth Schechter

Abstract

This paper introduces to the philosophical literature on personal identity a new candidate case of multiple persons in one body. *Plurals* are human beings who identify as multiple persons sharing a brain. I unpack the meaning of the plural identity claim and attempt to say something about its phenomenological basis. I argue that it makes sense to delineate plurals in terms of their shared identity, despite plurals' etiological diversity, and offer some possible explanations for the overlap between plural, trans, and autistic populations. The paper neither defends nor rejects the plural identity claim, but argues that, on the one hand, it is not clearly delusional, and, on the other hand, that there is a difficulty with trying to make sense of it from a third-person perspective.

1 Introduction

The two real-life hard cases that figure most heavily in the philosophical literature on personal identity are the split-brain phenomenon on the one hand and dissociative identity disorder (DID) on the other. Human beings from these populations are generally taken to be the best candidate cases of multiple persons in one body, because of a common feature the conditions share: systematic causal dissociations between ordinarily integrated, personal-level psychological states.

This paper will focus on and introduce to the philosophical community a population that substantially overlaps with but is distinct from the DID population. This population, which I will call *plurals,* raises the issue of personal identity not first and foremost because its members are subject to causal dissociations between personal-level psychological states but rather because they themselves explicitly identify as multiple people sharing one brain. Or, rather—speaking more carefully though admittedly more awkwardly—a plural is a human being out of whose mouth issues the identity claim (or whose brain produces the avowed belief whose content is expressed by the sentence), "I am one of multiple people in this brain." I call this the *plural identity claim*.

The population of plurals is unknown to most philosophers and poses several problems of understanding. First, I'll articulate the features of plural identity and describe the different etiologies of plurals, while arguing that despite their etiological diversity, it makes sense to delineate this population in terms of their shared identity. I will discuss possible explanations for the substantial overlap between plural, trans, and autistic populations.

Without arguing that the plural identity claim is true, I will try to characterize its meaning and phenomenological basis, and will defend it from two objections according to which the claim needn't be seriously investigated. The goal, throughout, is to sketch a kind of philosophical problem space around this population.

2 Preliminaries

My focus in this paper will be on human beings in the West who meet the criteria that I specify for being plural. There may be, in other cultures, human beings who would also say that their bodies are shared or inhabited somehow by multiple people, and such human beings may differ significantly from the human beings I discuss here. It would make sense to distinguish them, since human beings in very different cultures might be expected to possess very different concepts relevant to the self-belief specified in the doxastic criterion for being a plural, and this difference might in turn affect their experiences and ways of living.

The population of plurals focused on here overlaps substantially but not entirely with the population of human beings with dissociative identity disorder (DID). Human beings with DID are often called multiples, partly because the condition was once named "multiple personality disorder," but also because a striking feature of the condition is the experiencing of oneself as psychologically multiple in some way. It would be difficult to make clear what experiences of multiplicity are, and they are probably diverse; I will say more about the phenomenology of multiplicity at other points in the paper. Crucially, the self-beliefs of multiples are also diverse and only some multiples end up explicitly identifying as multiple persons.

I define being plural first and foremost doxastically: a plural is a human being who explicitly *believes* that there are in fact multiple persons sharing their brain. I call this belief *plural identity,* though some further elements must be specified to give its precise intended meaning. (Note that I am offering my own account of what it is to be a plural; plurals themselves arguably use the term "plural" synonymously with the term "system," which I define below.)

Some of the research drawn on in this paper is academic research of a traditional sort, including psychological and philosophical literature on DID. While much of that literature is relevant to plurals in some way, it does tend to lump together subjects who are very diverse with respect to their own self-conceptions. The core idea of this paper is that there is a philosophically, psychologically, and sociologically interesting population here properly delineated in terms of its members' self-conceptions. But there is very little academic literature on that population, so defined.

Fortunately, the paper is also able to draw on some sources that were not available to prior philosophical thinking about candidate cases of multiple persons in one body. Much of the philosophical literature on DID was written closer to the 1990s (Dennett and Humphrey 1989; Braude 1995; Hardcastle, Flanagan, and Institute 1999; Radden 1996; Sinnott-Armstrong and Behnke 2000), and since then, the internet exploded: many online resources and communities have sprung up around and for self-identified multiples and plurals, and these online sources offer important phenomenological and anthropological evidence.

In addition, I have had opportunities to speak with a number of plurals about their beliefs, lives, and experiences. These (recorded and transcribed) conversations do not qualify as scientific research per se, since I have no training in interview research. (I attempted something like semi-structured interviews but did not code them afterwards.) I also did not attempt to recruit a representative sample: the plurals I interviewed volunteered for the project after learning of it at one of a number of online communities for plurals, where an advertisement to participate in my research was posted by a plural who had contacted me after I published a popular piece on what it means to respect plural identities. Despite this haphazard method of recruitment, however, the plurals I interviewed were more diverse along certain dimensions—both age and race/ethnicity—than I had expected.[1]

Still, the present work does to some extent run ahead of the data, and it would be a happy outcome if it helped motivate further research. Towards that end, the paper introduces a number of concepts and distinctions useful in thinking about this new area of study.

Often, the language I employ in discussing plurals may seem to simply assume the metaphysics that plurals themselves endorse, according to which each plural human being is somehow associated with multiple persons. My intention in using that language, however, is not to beg any metaphysical questions, and in fact the paper does not argue that the plural identity claim is true. Indeed, many or most of the interesting philosophical (not to mention sociological and psychological) questions raised by the phenomenon of plural identity don't closely concern the metaphysics of plurals at all. One reason to employ the language that plurals use to describe themselves is that the plural identity claim will likely be challenging for many philosophers to comprehend, and I believe that employing language that assumes plurals' own perspectives makes it easier to understand what and who they claim to be. In addition, language that sounded more metaphysically neutral would often be hopelessly awkward. It is no surprise that plural communities are

---

1.   In this respect, they were consistent with the participants in Turell et al.'s (2023) interview research with transgender plurals: of the first fifteen who replied and met criteria in that study, about a third were white, about a third were mixed race/ethnicity, and of the remaining four, two were Asian and two were African American; their (bodily) ages meanwhile ranged from 18 to 38. That study similarly used community-based participatory research design.

themselves sources of terminology that enable their lives and their beliefs to be more easily understood and less awkwardly described. That said, there are also points at which I make terminological choices that are not (or not routinely) made by plurals themselves, since I need terminology that reflects and refers to the specifically *philosophically* relevant distinctions drawn in this paper. I will try to flag, as much as possible, what precisely I mean to commit myself to in utilizing one term or another.

3 Who and What Are Plurals?

Although the term "multiple" is usually used to refer to a human being with dissociative identity disorder (DID), I will instead take a multiple to be any human being who routinely has *experiences as of* there being other psychological beings present with them in their body or brain, regardless of whether or not they "have DID" (which itself could mean different things, depending on how closely one hews to current diagnostic criteria). In this paper, then, multiples are phenomenologically defined, while plurals are doxastically defined: plurals actually *believe* that there are multiple persons sharing their brain.

Most readers of this paper are, I assume, *singlets:* human beings who neither *believe* that they are multiple persons in one body nor are subject to "*experiences* of being or having, more than one individual within a single body" (Garrett 2023, emphasis added). So, as I will use the term, a singlet is someone who is neither plural nor multiple.[2]

I have defined multiples as human beings who *feel* as though they are psychologically multiple somehow. This feeling appears to be genuinely phenomenological, rather than just an inference drawn from their phenomenology: that is, while a multiple with DID may, for instance, find themselves behaving in ways they don't understand, perhaps (tenuously) supporting the inference that some of their actions aren't their own, multiples also *feel* as though something they are doing or something they are hearing in their mind is not really them but is rather someone or something else.[3]

This feeling of being multiple somehow is a striking feature of dissociative identity disorder. The most recent version of the Diagnostic and Statistical Manual for Mental Disorders—the DSM-V-TR—lists DID as one of several dissociative disorders, and gives as its first two and most distinctive diagnostic criteria for the condition:

---

2. Some plurals use the term "singlet" to also include to multiples that aren't plurals (Stronghold 2021); I'm not sure how common that usage is. Note that either way, within the plural community and the multiples community, being multiple and being a singlet are not considered to be exhaustive options; *medians* are at least one intermediary category. Medians however are not plurals as I define them, and I will not discuss them here.

3. Braude's (1995) analysis of multiplicity is still very helpful and addresses both its cognitive and phenomenological aspects.

A. Disruption of identity characterized by two or more distinct personality states… The disruption in identity involves marked discontinuity in sense of self and sense of agency, accompanied by related alterations in affect, behavior, consciousness, memory, perception, cognition, and/or sensory-motor functioning. These signs and symptoms may be observed by others or reported by the individual.

B. Recurrent gaps in the recall of everyday events, important personal information, and/or traumatic events that are inconsistent with ordinary forgetting. (American Psychiatric Association 2022)

Like those of many other disorders listed in the DSM, the diagnostic criteria for DID also require that the symptoms "cause clinically significant distress or impairment in social, occupational, or other important areas of functioning" (American Psychiatric Association 2022).

Criterion A's reference to "discontinuity in sense of self and sense of agency" may not clearly describe the felt sense of *multiple* psychological beings. This striking feature of DID is, however, targeted by some clinical measures, such as the "I have parts" items on Dell's (2006a) Multidimensional Inventory of Dissociation, and I will assume that multiples include all people with dissociative identity disorder. This assumption may not be correct and is not crucial in what follows. What is crucial is that plurals are a subset of multiples.

3.1 Plural Identity

It is pragmatically necessary to have a term to refer to the different beings that a plural believes are sharing their brain. The clinical term for such beings in someone with DID would be "alters," but plurals themselves often dislike this term for being "dehumanizing" (depersonalizing):

…they are not 'my alter.' No one is an 'alternate' to myself. We are a group of people. (The Blackbirds, n.d.)

A commonly used alternative is the term "headmates"—like "housemates," except sharing a head instead of a house. (The term "head" here is helpfully neutral between whether headmates are believed to share or experienced as sharing a brain, a skull, or perhaps even a mind in some way.) I will use the term "headmates" myself, although without meaning to beg the question of whether headmates are *entities*, as the term suggests, or whether they are instead more property-like. Plurals often refer to the collection of headmates associated with one body or brain as a *system,* and I will sometimes use this language as well. (Headmates are therefore sometimes called "systemmates" instead.) Note that while I use

the term "plural" to refer to a *human being* with a particular identity, the term "system" is slightly different, referring instead to the collection of headmates all associated with one particular plural.

The seemingly unitary criterion for being plural that I propose—believing that one is one of multiple persons in one's body or brain—involves a belief or network of beliefs that is in fact complex and that can itself be articulated in terms of five features.

The first feature of plural identity is that it is in part phenomenologically grounded: a plural doesn't just *believe* that there are multiple psychological beings sharing their brain but has experiences as of these multiple beings. A headmate may for instance experience their body as performing actions or "hear" thoughts in inner speech or be directly aware of emotional responses that register to them phenomenologically as *someone's* but also as *not-mine* or *not-me*.

The second feature of plural identity is that—unlike some multiples—plurals in fact *believe* that there are multiple such beings sharing their brain. Indeed, I am taking plural identity to be a fairly explicit self-belief, especially, as we will see, a belief that has implications for how they live their lives.

Third, plurals accept that the multiple beings whom they believe are sharing their brain are in fact *persons:* they believe that these beings—whom they call people—can have personalities, preferences, varying degrees of agency (including moral agency), self-consciousness, and, crucially, rights:

> We have a very strong commitment to operating collectively and look viewing each other [sic] as persons with equal dignity and deserving our existence. We make decisions by a process of consensus, which is not unanimity. (Turell et al. 2023, 5; quoting Jesse System)

But—and this is the fourth feature distinctive of plural identity—a plural does not believe that the overarching system or the human being as a whole is itself a person, except perhaps in a metaphorical sense. Contrary, then, to what some of the language used so far may have suggested, a plural human being does *not* make first-person singular statements to the effect that he is the *collection* of people inside him. That is, a plural does not say, "I am multiple people sharing one brain." Rather, a plural human being will say either, "I am one of multiple people sharing this brain" or "We are a group of people sharing a brain." Either way, a plural possesses one or more beliefs whose content is expressed in the first-personal sentence, "I am just one of multiple people here":

> Fundamental to properly understanding what I'm trying to convey is the proper placement of the "I." *I* call myself One. *I* am writing this letter. *I* am only myself; *I* have

one identity, one sense of self, one personality. Although I am conjoined inseparably from the other members of my group, when I am not on [sic?] front [not controlling behavior] my own "I" is no longer there; I am in a state like sleep; some other person now walks around in our body. Someone who has their own "I," their own internal narrative, their own wants and desires. When they step away from the front, yet another "I" will take their place…If someone can understand that essential fact at the heart of what multiplicity is, then phrases like "your other selves," or "when you were that other person," or "the other you" become obvious non-sequiturs. I don't have "other selves." *I* am never anyone but myself. (One Fox Faraday 2015, original emphasis)

Again, what a plural claims is that within their brain there are multiple headmates, each of whom thinks, *of that very headmate,* and not of the others, "I am *this* person." Analogously, I believe of my sister and I that *we* are multiple persons; I also believe, of Elizabeth Schechter, that *I* am *this* person; I certainly don't believe that *I* am *us*. A plural might express this fourth feature of plural identity by saying something like "I'm so-and-so, and I have three headmates, so there are four of us in our system."

This fourth feature is arguably entailed by the third: groups aren't literally persons, since they have some but not all necessary features of persons. (For instance, groups aren't sentient.) Therefore, to the extent that a plural believes that her headmates really are distinct and genuine persons, she won't also believe that *all* of them jointly constitute *one* literal person.[4]

It follows that earlier and future references to what a plural *says* or *feels* or *believes* are inevitably ambiguous. Sometimes it can mean simply that there is at least one headmate within that system who says (etc.) such-and-such. Sometimes it means that all of them jointly say or would say such-and-such, or that a majority would say so. In talking to a plural, it is not always clear which headmate is speaking, nor is it always clear whom is being spoken for. Indeed, this may not be clear to the very headmate speaking. Headmates often shift back and forth between the first-person singular and the first-person plural to indicate how widely within the system their perspective is shared. But sometimes a headmate may not know how widely shared is their perspective on something, much less exactly which other headmates share it. There is no way around this linguistic ambiguity in discussing what plurals think and say and so on, and this ambiguity should be kept in mind going forwards.

---

4. Some plurals think of the system itself as something with its own quasi-psychological properties, but, again, not exactly as a person. One headmate told me that their mental image of their system was as something like a giant gear somehow *enabling* interactions between headmates.

The ambiguity also pertains to the plural identity claim itself. So as not to beg any metaphysical questions, I have defined the claim in an awkward way: the claim ("I am one of multiple persons in this brain") issues from a human being, but if what plurals say about themselves is true, the human being is not the claim's proper maker. To plurals, the truth of the plural identity claim, for a particular human being, in fact requires that there are, within that human being somehow, at least two *potential* makers of that very claim. Within the plural community, plural identity may be understood as something closer to a collective identity: "We are multiple persons sharing this brain." If there is no *we*—if, contrary to the plural themselves, only one of those persons in fact exists—then plurals themselves would consider the identity to be mistaken. Indeed, plurals themselves seem to use the term "plural" as a synonym for "system": it applies only to a collection of systemmates, and if plurals as *I* define them are wrong and systemmates do not really exist —except as distinct property-clusters of one person—then there *are* no systems and no plurals, as *they* define them.[5]

I have not made the same terminological choice here and have instead taken a plural to be a human being who has a certain self-belief, whether or not that belief is true. (If the belief is false, then then it is the human being—or the sole person that human being constitutes—who believes it; if the belief is true, then the human being is merely associated with it somehow, its brain being the brain of the headmate who is its proper believer.) I have done that partly so as not to beg any metaphysical questions and also to leave open the possibility for an argument that a plural's self-belief itself grounds the truth of plural identity. That is, rather than defining "plural" in such a way that it is an open metaphysical question whether plurals exist (or whether instead some human beings just think that they do), I have defined "plural" such that plurals uncontroversially exist—they are human beings who identify as multiple persons sharing a brain—and this leaves space for the possibility that plural identity partly *grounds* the truth of the plural identity claim.

The fifth and final defining feature of plural identity is this: because each headmate thinks of themselves as a person sharing a brain or body with other people, each therefore conceives of their (that is, his or her or their) relations to the others as essentially *interpersonal*—rather than intrapersonal—in nature.

Consider the attitude of *liking someone*. This is first and foremost an attitude a person can have towards another. We do sometimes speak of liking or disliking oneself, but this usage is arguably metaphorical, the concept applying only partially in the reflexive

---

5. There is a difficulty here which I'm not sure how to get around: what about a system only one of whose headmates endorses the plural identity claim? (This is theoretically possible; I did for instance speak to one system of about ten headmates, one of whom insisted that they must all be parts of one person.) Relative to the way I've laid out the criterion for being a plural, such a human being *would* count as a plural, but I'm not sure plurals themselves would agree; perhaps this hypothetical edge case shows the limits of my approach.

case. For instance, I can't imagine liking someone without in any way enjoying their company, and I wouldn't know how to answer whether I enjoy my own company, unless this just means to ask how contented I feel during the times when I don't have company. In just the same way—from what I can tell—a headmate might be more readily able to say whether she likes one of her systemmates than whether she likes herself. Different headmates speak of liking or disliking, respecting or disparaging, cooperating and arguing and negotiating with each other. Conflict within a headmate (i.e., ambivalence) tends to be experienced differently from conflict between headmates. Systemmates feel gratitude and resentment towards each other—in a way that I don't feel gratitude towards the self-of-yesterday who packed today's healthy lunch (even if I am grateful that I packed lunch) or resentment towards the part-of-me-that's-capable-of-delayed-gratification for not "allowing" me to spend my whole paycheck (even if I resent that I have to economize). Some pairs of headmates are friends with each other; some pairs are not presently on speaking terms; some may be romantically and even sexually attracted to each other; others may deny that they feel attracted to each other only because of their own internalized homophobia (Riesman 2019)!

I have just offered five features that define plural identity or self-belief. This account is only provisional, however. In particular, I have set things up such that a phenomenological element is essential to having a plural identity in the relevant sense, but I am not confident that this is the correct choice. Certainly, a merely theoretical belief that one is one of multiple people in one's body is not enough to be a member of the population of interest: suppose that, in order to resolve a paradox of personal identity that I've been struggling with, I am led to conclude that there must already be multiple persons in my body even now, and that I am just one of them; I would nonetheless not have a plural identity of the sort of interest here. But it is unclear whether this is only or even primarily because my belief that I was one of multiple people in my body would not be grounded in phenomenology, or rather because I would lack further beliefs about these postulated multiple persons that plurals possess; for instance, even if I concluded, on a purely theoretical basis, that I was one of multiple people in this body, this needn't dispose me to try to communicate with them or even think that it was possible to relate to them in genuinely interpersonal ways.

Provisionally, however, I have chosen to delineate the relevant population in partly phenomenological terms—and to be fair, all the plurals I've spoken to have described atypical phenomenological experiences of their bodies and of their (brain's) thoughts, experiences, and/or behaviors. That phenomenology overlaps that described by the criteria for depersonalization/derealization disorder in the DSM. The phenomenology of depersonalization is characterized by "Experiences of unreality, detachment, or being an outside observer with respect to one's thoughts, feelings, sensations, body, or actions (e.g., per-

ceptual alterations, distorted sense of time, unreal or absent self, emotional and/or physical numbing)"; the phenomenology of derealization is characterized by "Experiences of unreality or detachment with respect to surroundings (e.g., individuals or objects are experienced as unreal, dreamlike, foggy, lifeless, or visually distorted)" (American Psychiatric Association 2022). Experiences of depersonalization and derealization are not so uncommon (see Michal et al. [2011] and Žikić, Ćirić, and Mitković [2009]). But commonly one experiences oneself as, as it were, detached from one's body or actions or even emotions —as though they were not one's own, perhaps, and thus in that derivative sense someone else's—without having any sense (or believing) that there is, simultaneously, within one's body, another being for who those actions and emotions feel like their own. Section 6 will say a little more about the phenomenology of multiplicity as a (partial) source of plural identity.

Plural identity itself is, however, first and foremost doxastic. Within this doxastic element, there will still be some diversity. I have required the belief to be an explicit or conscious one, for instance, but explicitness and consciousness arguably come in degrees; in some cases the belief could perhaps be closer to a background assumption structuring one's emotional life and decision-making, while in other cases it could be a very explicit theoretical belief, arrived at after a deliberative process of attempting to make sense of one's experiences; in other cases it is a kind of a sociopolitical identity, which forms the basis for organizing and activism alongside other plurals:

> The solution to medicalists in the plural community is much the same as the solution to similar attitudes in the trans community. Activism and visibility. Non-disordered, non-traumagenic, and mixed-origin systems must keep speaking out and sharing experiences. This is vital.[6] (RSpacefox 2021)

> [We] like the idea of like trying to spread awareness and education and all that. We love that and because we are very involved in disability justice, we think it's very important. (Turell et al. 2023, 7; quoting Finley System)

In short, I take plural identity to be a partially phenomenologically grounded belief whose content, expressed in the third-person singular, is that one is one of multiple people in one's body or brain, to whom one can relate in (many) ordinary interpersonal ways, and who are neither parts of one's own person nor collectively constitutive of a unitary (genuine) person of whom one is also partially constitutive.

---

6. This paper often pulls from blogs, online forums, Tweets, etc., that contain obvious typos. These have been fixed, for readability, when there is no concern that they will affect the meaning. For example, some posts said "they're" when they meant "their", etc.

3.2 Plural Etiology

I assume plurals include many human beings with dissociative identity disorder (DID), but not all or even necessarily most. Some multiples with DID experience themselves as being psychologically multiple but do not seem to believe that they are; many identify, in some ultimate sense, with all of their "parts"—or, if they really don't identify with them, they don't, at least, view them as *persons*. Note that this may not be because of, say, any differences in basic phenomenology between plurals and non-plurals with DID. The classic *clinical* perspective, after all, is that DID involves a single fragmented person, rather than a multiplicity of genuine people, and that healing consists of the progressive integration of this person into a psychic whole, and many human beings with DID have adopted this clinical perspective. This perspective is of course precisely what plurals reject:

> On the matter of being "whole": I'm already whole as a person, and the act of smashing the rest of us into an idealized single person wouldn't even work. We're all separate, with fully developed personalities and interests. We're individuals, and prefer to be treated as such. (Flynn 2011)

Many multiples with DID therefore are not plurals. Conversely, there are plurals who do not have DID. (See Figure 1.) Of those plurals who don't have DID, many once met diagnostic criteria but no longer do, while remaining multiple. They may cease to meet criteria because they no longer meet the distress/impairment criterion; on clinicians' parts, the judgment as to whether or not a multiple merits the diagnosis of DID will probably especially often concern whether the client's multiplicity *per se* is impairing them (see e.g., Vignettes 3 and 6 in Ribáry et al. [2017]). But plurals may also not meet diagnostic criteria because they no longer meet the amnesia criterion as the latter is framed, since multiple headmates may share their knowledge and experiences with each other (more on this below). Some plurals identify with the diagnosis to the extent that they believe that their system was produced by trauma—a major factor in the etiology of DID—but claim that they never strictly met diagnostic criteria. (At least, as they understand them; one can't always say what clinicians would have said, since some plurals have not come into contact with the psychiatric profession or have gone only to seek treatment for, say, one or more headmates' depression.)[7]

---

7. The paper assumes (although nothing really hinges on this) the traumagenic model of DID. The other best known model of DID is the sociocognitive model, which basically posits that DID is iatrogenic. It may be that there are two stages of DID, with the traumagenic model explaining the first stage and sociocognitive factors explaining the second stage, which might be called the "social stage" of the disorder, at which point what has previously been experienced as, say, voices, or inexplicable and foreign-feeling emotions, begin to be conceptualized as in some sense socially real individuals. The classic sociocognitive model does not

Of course, if one takes being a singlet, with a singlet's phenomenology, to be *normative,* then all multiples—including all plurals—will inevitably meet the first diagnostic criterion for DID. But without this assumption, some multiples will not meet the first criterion either. From the standpoint of plurals, in particular, the claim that they experience "discontinuity in [their] sense of self and sense of agency" (American Psychiatric Association 2022) just begs the question: headmate H1 may experience no discontinuity in her sense of self or agency while headmate H2 may experience no discontinuity in his sense of self or agency either. Of course, H1 doesn't have the sense of *being* H2 and doesn't feel as though H2's actions were *her* actions—but then, I don't have the sense of being my sister, and I don't feel as though her actions were my actions, and we don't take this to reflect discontinuities in my sense of self or agency because we just take it as given that my sister and I are different selves, different agents. But that is precisely what H1 and H2 claim to be.
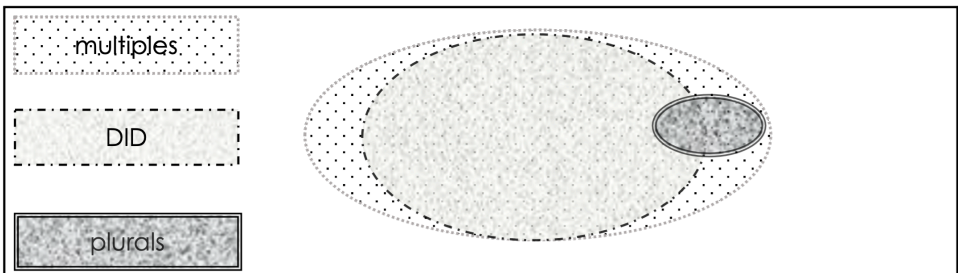
Figure 1: Three overlapping populations: multiples, DID, and plurals. Multiples include but are not limited to human beings with DID. I assume in the paper that all human beings with DID are multiples, but nothing hinges on that here. All plurals are multiples, but not all multiples are plurals. Some but not all plurals have DID. (Figure not drawn to scale.)

Importantly, there are also systems that don't have their origin in trauma to begin with. Some systems are intentionally created through so-called tulpamancy. Tulpamancy is a practice or set of practices undertaken with the intention of creating an autonomous sentient being "inside" (and of course using) one's brain; beings created in this way are called tulpas, and the people who created them are called tulpamancers. Tulpamancy has received a little popular media attention (Thompson 2014), but not much academic attention, although Vessière (2016) and Laursen (2019) are important exceptions. People engage in this practice for a diversity of reasons, ranging from simple curiosity to lone-

---

clearly distinguish between these two stages, however, whether because it identifies DID only with the second stage or whether because it believes that sociocognitive factors explain *both* stages. Note that the sociocognitive factors that the model appeals to are first and foremost therapist expectations. But it may be that there are other social, motivational, and conceptual sources of the development of the social stage, now that DID is more widely known.

liness and the perceived desirability of creating a companion that one can carry around inside oneself, in a sense.[8]

Then there are so-called "natural" or "endogenic" systems. Some claim that they were just always multiple people, without ever having experienced childhood trauma of the sort that is generally believed to be the precipitating factor for DID and without having intentionally and effortfully created headmates in the way that tulpamancers do; other natural systems say that while they have experienced such trauma—just as have many singlets—they were already multiple by that time. Natural systems' causal origins could perhaps just be some kind of neurobiological difference (or abnormality); alternatively, several systems I spoke to expressed the belief that *authors* may sometimes inadvertently create headmates in the process of vividly imagining fictional characters (see, on this note, Taylor, Hodges, and Kohányi [2003]). Note that this could be viewed either as inadvertent tulpamancy or—from the standpoint of a narrative account of the self—as just the same sort of process by which a singlet brain "creates" *one* person (Dennett 1992).

Systems, then, can have one of at least two and possibly three causal origins (see Table 1. Some are traumagenic, that is, caused by trauma and trauma-induced dissociation; these are the systems most likely to meet diagnostic criteria for DID, especially Criterion B. Some are intentionally created; these are what I am calling tulpagenic systems. Finally, there may be "natural" systems, the product neither of intentional effort nor of trauma and trauma-induced dissociation. However, since being a natural system is something of a diagnosis of exclusion, the status of this third type of etiology is less clear, and I won't discuss natural systems in what follows.

On the face of it, tulpagenic and traumagenic systems are very different etiologically, though the extent of this difference is difficult to resolve at present, when there remains ongoing debate about the nature of dissociation and the mechanisms of traumagenic DID. (Compare, for instance, Nijenhuis and Van Der Hart [2011] to critical responses to that article by Butler [2011], and by Dell [2011].) Certainly it may turn out that there are mechanisms common to both and even to non-pathological forms of dissociation. There can be other etiological overlaps as well, as when a multiple with DID intentionally creates just one of their headmates; additionally, both tulpagenic and traumagenic headmates may be modeled on fictional characters ("fictives") or on real people known to the plural ("introjects") or on versions of one's own (real or ideal) self (this list of options is not exhaustive)–the choice of model simply being more intentional in the case of tulpagenic systems. Still, systems do appear to have (at least) two broadly different etiologies:

---

8.   Many people—including tulpagenic systems themselves—have expressed some discomfort with the use of the terms "tulpa," "tulpamancy," and so on, and have in some cases expressed negative attitudes towards Western practices of tulpamancy themselves, due to concerns about cultural appropriation (Mikles and Laycock [2015] discuss some of the relevant history and evolution in Western understanding of these originally Buddhist ideas). Convergence around an alternative set of terms has not yet emerged, however.

trauma-induced causal dissociation in the one case and in the other certain kinds of intentional imaginative and meditative practices.

System Type	Origin	DID Criteria	Other Characteristics
Traumagenic	Childhood trauma and trauma-induced dissociation	Likely to meet or have at one time met Criterion A, B, and/or C.	May have lived as multiples for a long time before identifying as plural. System may be very large.
Tulpagenic	Intentional effort to create one or more headmates	Likely does not meet criterion B or C; meets criterion A ("discontinuity in sense of self and sense of agency") only if being a singlet is normative.	Clear "original headmate" (tulpamancer). Since original headmate attempted to create other headmates (tulpas) with particular traits, headmates may be very similar. Less "dissociated" especially in distressing or impairing ways.
Natural	Unknown	Unknown	

Table 1: Types and origins of systems.

In terms of their internal system dynamics, tulpagenic and traumagenic systems tend to operate differently, with headmates in tulpagenic systems being much more aware of each other's thoughts and experiences and actions than they are in traumagenic systems. This is natural, since trauma is one cause of dissociation.

Nonetheless, when tulpamancy is successful, the tulpamancer experiences their tulpa or tulpas as being autonomous beings, just as occurs in traumagenic systems: so, although the tulpamancer will be aware of their tulpa's (say) actions, they will *feel* as though they (the tulpamancer) are not the agent of those actions. Phenomenologically, then, all plurals seem to share something.

There are interesting psychiatric connections between DID and non-traumagenic plurals: in particular, non-traumagenic plurals may have implications for the best methods of treating people with DID. DID is still commonly thought of as disordered not just because of amnesic and other (functionally impairing) dissociative symptoms, but also because it involves multiplicity per se. Indeed, some sort of fusing of the multitude of alters (headmates) into a single self is still widely viewed as the final stage of treatment, the sine qua non for full recovery. (Although I gather, anecdotally, that clinicians increasingly let patients take the lead in deciding when and even whether to ever pursue fusion.) The fact is, as Laursen (2019) says, that "Since the majority of people claim to have a single identity, there is a common cultural and psychiatric assumption that this is the most healthy, functional way of being" with "multiple identities tend[ing] to be viewed as problematic,

associated with pathological diagnoses" (172). But since tulpagenic plurals do not meet the diagnostic criterion for DID (nor perhaps for any mental disorder), they may show that being multiple, and even having a plural identity, is not in and of itself unhealthy.[9]

One might object that, since the DSM's criterion D for dissociative identity disorder requires that the disruption of identity referred to in criterion A not be a part of a "broadly accepted cultural or religious practice", tulpamancy cannot be relevant to evaluating the health of multiciplicity in DID—at least if one also thought that tulpamancy is such a practice; Laursen (2019), for one, simply states that it is. Admittedly, it might sound surprising to hear tulpamancy referred to as "broadly accepted," given that most readers won't even have heard of it—but on the other hand, who gets to say to which culture a tulpamancer belongs? The online tulpamancy community might itself be said to be one of many micro-communities and micro-cultures that have proliferated online.

But rather than showing that tulpagenic systems are irrelevant to evaluating the healthiness of multiplicity in DID, Criterion D's very existence might instead be said to frankly *acknowledge* that multiplicity *per se* is not necessarily disordered. And while it is true (I am assuming) that multiplicity in DID is caused by trauma, this too does not clearly impugn multiplicity in DID, since it cannot be assumed that every spontaneous response to trauma is itself unhealthy. These sorts of considerations at least raise the question of whether fusion needs to be even an ultimate or ideal outcome of DID treatment.

3.3 Plurals and Plural Identity

As just explained, there are differences between tulpagenic and traumagenic systems with respect to etiology and psychodynamics. From the standpoint of this paper, however, what distinguishes tulpagenic and traumagenic plurals is less important than what they share: endorsement of the plural identity claim. Traumagenic systems seem to be much more dissociated, and robust and systematic causal dissociations between personal-

---

9. This is not to say that tulpagenic plurals enjoy perfect mental health. Many self-report one or more mental health diagnoses, especially depression, anxiety, and ADHD (e.g., Shinyuu [2015]); indeed, a frequently cited reason for trying to create a tulpa is to ease feelings of loneliness that may be more prevalent in the mentally unwell; others state that they attempted to create a tulpa specifically to help deal with a mental health crisis (BenitoFlakes_ 2021). But loneliness, depression, anxiety, and ADHD are also common in the general population (and are higher in young people who are, unsurprisingly, also the ones most cognizant of and interested in tulpamancy in the first place). It is an interesting question whether the creation of one or more tulpas can be a healthy strategy for the depressed (or the merely lonely). Talking to a tulpa is a lot like "talking to God", and "talking to God" *can* be healthy (Zarzycka and Krok 2021). One might in fact worry that it works too well, disincentivizing a tulpamancer from seeking relationships with other *human beings*. But if the human mind truly is capable of seeing tulpas as people—and certainly if tulpas *are* people—then it's not clear that the distinction between socializing with people inside versus outside the system matters, from a mental health perspective. In any event, the important point is that tulpagenic systems' mental health problems may well *precede* (and perhaps contribute to) those systems' creation, rather than their systemhood *causing* (or constituting) such problems.

level mental states might themselves be argued to be the basis for a claim of multiple personhood. Self-identity, however, is another interesting potential basis, and I think there is some intuitive force behind the idea that it matters for personhood. Indeed, Vessière (2016) notes that in approving his interview research with tulpagenic systems—in which headmates are highly co-conscious—the review board "was concerned with the anonymity and protection of tulpa persons, as well as that of their hosts" (68).

Because *I* have delineated the class of plurals in terms of their explicit self-conception, plurals—again, as I define them—are not a natural kind, any more so than are, say, atheists. There may well be natural kinds relevant to the phenomenon of plural identity; dissociative identity disorder could be a natural kind, or it could turn out that there is a specific and scientifically explicable type of dissociative phenomenology systematically experienced by all and only plurals. It would also be possible to delineate the population differently, as the population of whom the plural identity claim were *true,* whether or not its members believed it: in that case, it could perhaps be debated whether members of that population (were there to be any) constituted a natural kind (the issue would partly depend on whether persons are a natural kind just in general).

But there are defensible reasons for a delineation based on self-conception. Many of the interesting philosophical, psychological, and sociological questions raised by plurals depend directly upon their identities. For instance, in what ways does conceiving of oneself as one of multiple persons sharing one's brain allow one to live differently, to relate to experiences of inner speech differently? Do plural identities themselves—and not the correctness of those identities—have ethical implications? What is the process by which learning about and hearing from other plurals comes to transform a human being's interpretation of their phenomenology, or even to transform that phenomenology itself? What sorts of practices or social contexts promote the stability of a plural identity, and are there practices or social contexts that make it likely that someone will lose that identity? Do plurals themselves have responses to the obvious philosophical arguments against the possibility of plural personhood? Is there a connection between the rise of tulpagenic plurals and the growing problem of social isolation (US Surgeon General 2023)?

There has been a little movement, recently, towards recognizing plurals, regardless of etiology, as a population. Turell et al. (2023) have published a paper on the experiences of transgender plurals of different etiologies. Christensen (2022) has a helpful paper on what she calls the *culture* of plurality, and although the paper explicitly focuses on DID plurals, she does (somewhat obliquely) mention tulpagenic plurals as well (2022, 3). Ribáry et al. 2017 present interview research with systems that seem to be uniquely traumagenic, but the authors define the population of interest in terms of their explicit self-identity (an identity that non-traumagenic systems may share, though Ribáry et al. [2017] recruited

self-identified *multiples* rather than *systems* or *plurals*). Laursen's (2019) article on tulpa-mancy helpfully relates tulpagenic plurals to plurals with DID.

It should be acknowledged that making the case for grouping some traumagenic and tulpagenic systems together on the basis of their explicit self-identity risks creating the impression of greater harmony between those systems than in fact exists. It is easy to find, online, groups of traumagenic plurals that deny the reality of non-traumagenic systems: from their standpoint, self-identified natural or intentional (tulpagenic) systems are in fact either unwittingly traumagenic or else mere appropriators who are not genuine systems at all. Against them stand non-traumagenic systems (and some allies) who decry their exclusion and accuse the former group of gatekeeping. The contours of this conflict —familiar from other contexts—are themselves interesting, but I set it aside here to focus on what unites all human beings who *claim* to be one of multiple persons in one head, regardless of their attitudes towards other such human beings.

It's natural to wonder the size of the population of plurals, but I do not know the answer. I know of one interview study of self-identified *multiples*, which estimated from online sources that there are "200-300 individuals who participate in these forums and believe they are multiple" (Ribáry et al. 2017). This estimate is surely too low, though, be-cause the authors used only the search terms "multiplicity" and "multiple system," whereas at the time, on Twitter (now X) and Tumblr, terms like "collective" and #pluralgang were used at least as frequently. The term "multiples" also seems to me to be more strongly associated with DID specifically, and indeed the systems discussed in that paper all seem to be traumagenic. One can't judge prevalence just by looking at the number of posting members of online communities for plurals, meanwhile, since it is common for different systemmates to each have their own profile and create their own posts. At the same time, whatever the membership of these online communities, there are presumably many more plurals than participate in them, and—as Christensen (2022) argues—their numbers are likely to grow.

Whatever their numbers, however, the mere existence of online communities for plu-rals is of sociocultural interest. Popular media articles have been written (Riesman 2019; Thompson 2014); Plural Pride events have been organized. Plural activists have attempted to raise awareness and have requested a kind of social recognition *as* plurals. Liz Fong-Jones, a former Google employee (and well known enough to have been written about in prominent media articles, e.g. Fried [2019]) for a time self-identified as plural on Twitter (now X, which Fong-Jones has left). Fong-Jones also appears to have had a hand in writing a document called the *Plural Playbook* (Batman and Irene, n.d.), which is also publicly available online, that was supposedly distributed several years ago at Google, to introduce managers and employees to plurality and offer tips on how to respond when someone "comes out" as plural.

4 Overlaps with Other Populations

While I am not in a position to quantify the extent of the following overlaps, I have observed (and this is confirmed by Christensen [2022]) that the population of plurals overlaps significantly with two other populations: the trans population and the autistic population.[10] Presumably, singlets make up the large majority of transgender people and of autistic people. But among the plurals I've spoken to, a large majority of them had either received an autism diagnosis or had self-identified as autistic, and almost all of them had at least one headmate whose gender identity was not that associated with the gender their body was assigned at birth. Here I want to say something about the most obvious possible explanations for these overlaps.

4.1 Transgender and Plural

For the purposes of this section, I'll take a transgender headmate to be any headmate identifying as a gender other than the one associated with the sex their system's body was assigned at birth. This is an awkward way of speaking—we assign both sex and gender to human beings, not to their bodies—but of course headmates are not animals (and neither headmates nor systems exist at birth anyway, presumably). Note that such a headmate may not actually identify as specifically transgender, nor may their system; the relationship between transgender identity and headmates' gender identities is complex (see Turell et al. [2023]). But I will refer to a headmate who identifies as a man in system whose body was assigned female at birth as a transgender headmate (and *mutatis mutandis* for a headmate who identifies as a woman).

There is an obvious conceptual link between the person making a trans identity claim and the person making the plural identity claim: both believe that certain basic facts of embodiment commonly thought to *determine* aspects of personal identity are actually *not* so determinative. Until extremely recently in the West, one's gender was thought to be fully determined by physical facts about one's body, and that is still how everyone (or almost everyone) thinks about the *numerosity* of persons. It seems possible that once one relaxes the constraints between embodiment and identity in one of these cases, it makes it easier to relax them in the other case as well. (Something like this emerges as a theme

---

10. The autistic plurals I spoke to pretty much all rejected a disorder view of autism. Some of them therefore also rejected language suggesting that autism is a *condition*—the language, that is, of a *person with autism* —and instead conceptualized autism just as another way of being, using the language of *autistic people* instead. I use the same language in this paper.

of recent interview research with systems that identify as transgender in some way [Turell et al. 2023].)

The plurals I spoke to suggested a different explanation for the high occurrence transgender headmates, which is the association between being transgender (in a cis-centric world) and dissociation. An assigned-female-at-birth child who believes not just that their thoughts that they are a boy are factually incorrect but also that there is something wrong with the very fact of having such thoughts to begin with, will literally attempt to dissociate from those thoughts when they occur. And feelings of gender dysphoria are themselves so uncomfortable that they, too, create a temptation to dissociate; indeed, one could argue that gender dysphoria *is* a kind of dissociation—a feeling of alienation from one's body. Dissociation is both an aspect and a cause of multiplicity (which, given the right conceptual framework, becomes plural identity). Some of the transgender headmates I spoke to basically said that they tried *not* to experience or to think about their bodies—and again, it is the unity of the body, rather than the mind, that is ordinarily thought to be determinative of one's unicity as a person.

Another possible theory of the overlap may occur to some readers. Suppose that one is assigned male at birth but in some perhaps inchoate way thinks of oneself as or wants to be a girl, but suppose that one distances oneself from or actively disavows these thoughts and desires. Mightn't one then construct a headmate of the gender with which one inchoately identifies—as a kind of wish fulfillment, an ideal self—but while denying that the image so constructed was in fact of *one's own* self? I asked this question of two tulpagenic systems whose bodies were assigned male at birth. In each case, the tulpamancer initially identified (albeit not quite comfortably) as a boy or man; each then intentionally created a female-identified tulpa; each tulpamancer then at some point ceased identifying as a man (in one case coming to identify as a woman and in the other case as nonbinary). Both conceded that their (years) earlier decision to create a female tulpa might have had *something* to do with their own uncertain or confused gender identities. Still, they thought the connection was fairly indirect: their gender was something they were at least dimly aware of struggling with and of wanting to talk about, and they naturally wanted to have that conversation with someone gentle and accepting, and due to (they admitted) their own gender stereotypes, when they pictured someone gentle and accepting, they pictured a woman. But, they pointed out, the gradual shifts in their own gender identities did not make their female Tulpas feel less "necessary" or less *real*. So it wasn't as though approaching their "ideal selves" led to *merging* with their created companions or made those companions fade away.

A final obvious connection between transgender and plural populations is the internet. Transgender youth who feel isolated or aberrant at school or in other "real life" contexts may turn to online communities for social support and understanding, and *online*

is where one is likely to encounter plural communities as well. (It's where I learned of them.) Moreover, trans communities online may be more accepting of plural identities; indeed several transgender plurals I spoke to said that trans communities are both more likely to give weight to plurals' own plural identity claims out of deference to people's first-person epistemic authority but also—even when they don't necessarily believe those identity claims—more likely to treat them respectfully.

4.2 Autistic and Plural

It seems likely that an explanation of the overlap between the autistic and the plural community will need to refer to two different factors: phenomenological and other psychofunctional differences on the one hand and then a variety of social and doxastic or intellectual factors on the other hand. Indeed, there are many potential factors here deserving investigation.[11]

Some of the same explanations as above could be applied to explain why a disproportionate number of plurals are also autistic. Several of the autistic plurals I spoke to said that being autistic in an allistic (i.e., non-autistic) world is itself often traumatic—the traumas ranging from being misunderstood, to being uncomfortable in a physical and social environment not designed for people like oneself, to being actively mistreated, perhaps even by close caregivers—and trauma is a cause of dissociation. (See Reuben, Stanzione, and Singleton [2021] on autism and trauma.) Autistic people are also more prone to "sensory overload," and this itself can prompt dissociation as a coping mechanism.

There are potentially neurophysiological factors as well, involving the higher rate at which autistic people experience abnormal phenomenologies (see Ribolsi et al. [2022] for review). One thing to note is that autistic people may simply be disproportionately prone to experiences of *thought insertion*. To have an experience of thought insertion, with respect to a particular thought, is to lack what is called in the literature on the phenomenology of the self, a *sense of ownership* for that thought. (The liberature distinguishes the sense of ownership for a thought from the sense of agency for a thought; see e.g., Martin and Pacherie [2013].) The literature on the phenomenology of ownership (and on the phenomenology of agency) make clear that such experiences are the products of *inferences* (albeit often only at a subpersonal level) made by a mindreading system. And one of the ways autism manifests is as differences in mindreading ability or performance.[12]

---

11. It's worth noting that in the autistic systems I encountered, headmates believed that every headmate in their system was autistic. They viewed autism a something like a fundamental processing style that a brain either does or doesn't have, and thus as something all systemmates will necessarily share.

12. Some autistic people believe that what the literature calls "social difficulties" or "theory of mind deficits" in autistics should properly be called something like "social difficulties with allistics" or "theory of allistic

Typically, theory of mind differences in autism are discussed as deficits in mindreading others. But many researchers have claimed that people with autism are impaired at mindreading themselves, too. (Though, to be fair, much of the literature investigating deficits in psychological self-knowledge concerns something closer to an over-propensity to attribute intentions and beliefs to oneself, rather than an under-propensity to recognize one's mental states as one's own; see, e.g., Williams [2010] for review.) If such deficits are real, it is plausible that they could contribute to an increased tendency to lose the sense of ownership and the sense of agency. Arnaud (2020) meanwhile argues that self-consciousness simply operates differently, in a rather global sense, in autistic versus in allistic (non-autistic) people, with autistic people having (among other things) a greater tendency towards *third-personal* routes to self-knowledge (e.g., via observing their own behavior) and greater difficulties shifting between first- and third-person perspectives on themselves.

Returning to social and cognitive factors, several autistic plurals I spoke to in one way or another referred to how autistic people relate to social constructs. Allistic people, they said, tend to learn social constructs so effortlessly and intuitively, and often at so young an age, that they may have trouble even recognizing them *as* social constructs. In contrast, they said, for autistic people, learning about constructs like gender takes effortful and explicit thought—and once explicit thought is introduced into the process, there is the possibility of challenging the construct itself. And the *unitary person* is *also* a social construct: as one headmate said, "When the thing that you're doing explicitly is constructing an identity, that gives you a lot more opportunities to do weird things with it like construct two identities in parallel as opposed to just something that your brain does automatically, that you don't think about."[13]

---

mind deficits": allistics (i.e., non-autistics) have just as much difficulty understanding autistics, they say, as autistics have understanding allistics; autistic mindreading looks like a *deficit* only because allistic minds are so much more numerous. (This perspective would however be challenged were it in fact the case that autistic people are also less accurate at judging their own mental states.) Really, the only claim that would be needed to suggest a basis for an increased tendency towards experiences of disownership is the modest one that autistic people tend to apply theory of mind concepts somewhat differently than do allistic (non-autistic) people. There are cognitive psychological accounts of autism that make much stronger claims, such as that the core deficit in autism (such accounts simply assume a deficit picture of autism) is a theory of mind deficit. But the core feature of autism could operate at a much lower level—for instance, some sort of basic difference in perceptual processing—and still have downstream effects on the acquisition and/or exercise of mindreading abilities.

13. A couple of autistic plurals I spoke to also commented that autistic people are less motivated to try to be *normal* with respect to their identities and self-presentation and suggested that as a result, autistic people are also less averse to giving atypical answers to identity questions once they are explicitly raised. I do not know whether that empirical claim about the drive for normalcy is correct, though one paper (Chevallier et al. 2012) argues that while the strength of primary attachments (to parents, children) and the drive for sexual and romantic partnerships are basically normal in autistic people, desire for *social affiliation* is significantly reduced. The desire for social affiliation is itself multifaceted, but one aspect of it is impression management: acting so as to appear likable, attractive, competent, and so on; the authors cite a number

It should also be noted that autistic people often struggle with feelings of social isolation and loneliness (Schiltz et al. 2021), and such feelings may motivate creating headmates. Most of the plurals I spoke to said that conversations and relations between headmates are so similar to ordinary interpersonal conversations and relationships that it is difficult to feel *lonely* as a member of a system. And in fact, one frequently mentioned reason for pursuing tulpamancy is something like loneliness (Shinyuu 2015) and the perceived desirability of sharing one's day-to-day life and innermost thoughts with another being or person specifically created for affinity with oneself, for the rest of one's life.

There is in fact a kind of "hyper-individualism" that's interesting about the phenomenon of tulpamancy, where the society-wide problem of increasing social isolation and alienation is dealt with not by *reaching out* to the community of other human beings but rather by creating community within oneself. And it is possible that this sort of solution is more tempting to autistic people, in part because the challenge of connecting with (majority allistic) human beings—great as it is for the allistic population, at this point—can be even greater for them.

5 The Basis of the Plural Identity Claim

This section concerns the basis of plural identity and some challenges to making sense of the plural identity claim.

Identities are a subset of one's beliefs about oneself: those that provide a sense of meaning, purpose, or affiliation, and that offer constraints on action. Self-beliefs that are not part of my identity do not provide constraints on action in the same way; for instance, I believe that I am a Honda driver (both of our cars are Hondas), but only if this were a part of my *identity* would my being a Honda driver count for me as a reason not to buy a Toyota.

Someone's identity, so defined, is a subjective psychological feature; it is not, then, the objective notion of identity in which metaphysicians are interested, which I will call *metaphysical identity*. If animalism offered the correct metaphysical account of the identity of persons, then I would be an animal regardless of whether or not I identified as (or even believed that I was) an animal. After all, animalists believe that those who identify as immortal souls are also animals.

---

of papers suggesting that autistic individuals do far less of this. Of course, one might think that this is not due to lesser motivation to be accepted by others but lesser knowledge of *how* to be accepted by others. But the authors argue for a specifically motivational factor, drawing in part on neurophysiological research. It should be noted that the nature of social desires in autistic people is subtle and controversial (Jaswal and Akhtar 2019).

Crucially, good answers to questions of metaphysical identity can be bad answers to questions of *psychological identity*, and vice versa, since the criteria for goodness are so different. (Presumably answers to metaphysical identity questions are good if they're true; truth is at least not sufficient for goodness in the case of answers to psychological identity questions, and it might be doubted that truth is even necessary.) Many philosophers who endorse animalism probably don't find the proposition that they are animals to be great sources of meaning or purpose, for instance.

Most psychological identities are neutral on issues of metaphysical identity. Someone who identifies strongly as an academic isn't committing themselves to an implausible metaphysics of persons; they think that it is *they themselves* who (for example) "wouldn't know who they were" if they didn't have an academic position.

Sometimes, however, people's psychological identities do make metaphysical commitments. I have a New Age friend for whom it is absolutely central to her identity, including her life's work, that she is a reincarnated soul. It wouldn't surprise me if there were vegans who identified as animals in a way that mere animalists (typically) do not. A part of a plural's (or a headmate's) psychological identity, too, seems to be a proposition about her metaphysical identity conditions. If this is the right way of interpreting the plural identity claim—as a metaphysical claim—then it is naturally inconsistent with animalism.

The animalist Eric Olson once argued that if the diencephalic conjoined twins Brittany and Abigail Hensel are parts of one animal (as there is at least some reason to think), then they are in fact parts of one person (Olson 2014). In that case, he said, if one of their heads and brains—say, Brittany's—were destroyed, no *person* would have thereby ceased to exist, so long as the other brain (Abigail's) continued to function. Of course, the Hensel twins' parents would mourn *as though* they had lost a daughter, and it would be inappropriate to attempt to console them by saying, "I'm so sorry for what your daughter has lost, but at least no one died." Nonetheless, this interpersonally inept remark would be *metaphysically* correct, on Olson's account. Such implications have led me to believe that even if animalists are somehow correct about the metaphysics of the thing, they're not talking about what many of us mean when we talk about persons.[14]

For this reason, I believe that some version of psychologism offers the best account of the concept of personal identity that I am interested in. But animalism merely forbids plural personhood; psychologism does not show that it is nomologically possible, much less that it is actual in the case of plurals. And in fact, even if singlets all agreed that it is

---

14. Olson might respond that my confidence that Brittany and Abigail are in fact distinct person is explained by my believing that they are distinct animals. But in fact my own reasonably confident view is that it is metaphysically indeterminate how many animals Brittany and Abigail are, since they are biologically quite intermediate between being one human animal and two. It is not metaphysically indeterminate how many persons they are, though; they are two persons—using what I think is the ordinary (but also ethically primary) notion of a person.

theoretically possible for a human being to be multiple persons, there is something puzzling about *plurals'* own claims to multiple personhood. This is because plurals reject—as a description of multiplicity—the stereotypic picture of multiplicity inspired by DID. But it is this stereotypic picture that suggests criteria of individuation for persons that has up until now allowed singlets to make sense of the idea of multiple persons in one body.

5.1 What Individuation Criteria?

The stereotypic picture of multiplicity has two main elements, each of which bears some resemblance to one of the first two diagnostic criteria for DID, though in exaggerated form. The first element of the stereotypic picture is that distinct headmates have starkly different personalities, values, manners of social interaction, and so forth (see, for instance, Tye [2003])—and, indeed, that these stark differences between headmates are overtly manifest in behavior. Each of these distinct headmates is—according to the stereotypic picture—unusually one-dimensional in its personality and predispositions, with each headmate representing a different dimension; for instance, one headmate might be the "angry one" and another the "bubbly one." One might doubt (and many philosophers have doubted) that such one-dimensional types could truly be persons, rather than mere aspects of a single rich three-dimensional person. But at least this feature of the stereotypic picture suggests a clear criterion of individuation that might be used to argue for the multiple persons conclusion.

The second element of the stereotypic picture of multiplicity is that each headmate is a distinct island of consciousness and memory in a sea of amnesia: different headmates have different memories and personal knowledge, each being wholly unaware of the other's experiences and actions, except perhaps very indirectly (e.g., finding evidence on one's credit card statement of a purchase made by another headmate). This feature suggests a neo-Lockean criterion of individuation that might be used to argue for the multiple persons claim: one looks for continuity in (especially episodic) long-term memory—itself suggesting continuity of first-person conscious experience—casting discontinuities as the boundaries between different persons or selves.

The problem is that plurals often claim that the stereotypic picture is a picture of *dissociative identity disorder*, specifically, and an exaggerated one even then: plurals without DID and even plurals with DID often deny that it characterizes their own multiplicity. They deny, for instance, that their headmates are always of neat and distinctly defined personality, emotional, or behavioral types:

One of the beliefs perpetuated by early therapists and MPD/DID literature was that, since all selves were assumed to be a "part" of an original person, each of them was only capable of a single function or emotion—the angry one, the scared one, the seductive one, etc. There are people who *do* mistake having different aspects of their self for having separate selves, or name their moods and decide them to be different personalities. However, we're certain we're not doing anything like that. Why? Well, all of us have full ranges of emotion, and don't identify in particular with one emotion—we're all capable of being happy, upset, scared, angry, etc. We don't switch every time we start feeling a particular emotion. (Amorpha Household, n.d.)

Again, even multiples with DID may deny that their headmates are radically psychologically different from each other:

Nothing with DID is ever black and white and just like you and I might have things in common, there are plenty of things I have in common with everyone I share my brain with. (Callum 2019)

It is not just plurals and multiples themselves who reject this aspect of the stereotypic picture of multiplicity; DID clinicians and researchers have noted that different headmates (clinically, "alters") may have much in common, as Kluft (one of the best known DID clinicians) does when he notes that "Alters may pass for the host or be copies of the host" (2006, 297); indeed, Kluft goes so far as to claim that possessing "alters" (his term) that are "quite similar" or even "isomorphic" is the "true paradigmatic expression" of dissociative identity disorder (1991, 611). Clinicians and researchers thus now often emphasize that dissociative identity disorder is at base a *phenomenological* condition (see e.g, Gleaves [1996, 44], as well as Dell [2006b]).

Plurals (with and certainly without DID) may also deny that their headmates are mutually amnesic, in either of two senses. First, two headmates may be mutually co-conscious, both experiencing everything at the same time. Thus while two headmates may report having had different (token) experiences of each other's actions, they also both report having experienced and now remembering all the same actions. (It's worth noting that some research into amnesia in DID has suggested that it is a deficit in "meta-memory" (Huntjens et al. 2006, 862)—that is, *knowledge* of memory—rather than in memory itself (see e.g., Marsh et al. [2021]; Kindt and Van Den Hout [2003].) Second, even if two headmates are never co-conscious with each other, they may share knowledge of each other's experiences and actions by "communicating" with each other after the fact. (Many plurals report that their different headmates can "talk to" each other, in inner speech; others speak out loud to each other or write notes; other headmates seem simply to *know*,

propositionally, what the others have experienced.) Indeed, even traumagenic plurals may not meet Criterion B strictly stated:

> It is possible for multiple groups to have continuity of consciousness between persons – a memory as good as anyone else's. In a responsible, healthy system, if something important happens, people [i.e., headmates] will be told one way or another, even in systems with little or no co-consciousness. If we need to remember something, we will ask other people [i.e., other headmates] about it and obtain that information. (The Blackbirds, n.d.)

So, plurals reject the stereotypic picture of multiplicity for mischaracterizing the nature of the relationships between headmates. This rejection is, again, echoed by some contemporary clinicians and researchers working with DID patients. But this rejection makes it difficult to understand the meaning of the plural identity claim or the confidence in which it's held.

Using the stereotypic picture, it is at least clear what the criteria for individuating persons in a DID subject would be, even if we don't think those criteria are very good: either a sort of criterion of radical psychological difference, or an amnesia criterion, or both. Note that these are, effectively, third-personal criteria: extensive psychiatric test or observation could reveal whether or not they were met. But the multiplicity that plurals describe might well be basically *invisible* to second and third parties. This is not because there are *no* psychobehavioral differences between different headmates. It's because if we don't have any non-subjective grounds for positing them in the first place—no grounds, that is, other than first-person report—then there is no great obstacle to seeing multiple "headmates" merely as multiple features of *one* psychological being that make that being rich or nuanced or—at worst—simply confusing in some way.

In one sense, this isn't a surprising result, if multiplicity is in fact first and foremost a phenomenological condition, as Dell (2006b) argues. Still, most singlets don't have the phenomenology of multiplicity, and this presents a serious obstacle to their (our) understanding a plural's claim that there are multiple people sharing their brain. What criteria of individuation are *plurals* using? Not a spatial or bodily criterion, obviously, and not a (straightforward) criterion of psychological *difference*, either.

One possibility is that plurals mean to individuate persons on the basis of their consciousnesses. Unfortunately, there are longstanding puzzles about how to individuate streams or centers of consciousness (Famously, for instance, people are able to conceive of the continuity of either their neural or their other psychological properties *without* the continuity of their very stream of consciousness—of their own conscious subject, as opposed to a subject merely psychologically identical to it.) Worse yet, plurals don't

even seem to consistently mean to individuate headmates in terms of their streams of consciousness. For in many cases, they say, the brain gives rise to a single stream of consciousness—at least in the sense that there is never more than one such stream at any moment. Yet this consciousness is *colored,* at a given time, by the identity of the person whose consciousness it "is" at that time: so first, the world is experienced as Sarah experiences it, and then the world is experienced as Miguel experiences it—but (they report) it is less like each of them has their own stream of consciousness and more like first Sarah is subject to that stream and then Miguel to that same stream.

Plurals seem to use discontinuities in their *sense* of self and agency—discontinuities for which I don't think there is, currently, any adequate explanation—to draw the boundaries between different consciousnesses.[15] But this is still problematic. A plural claims that *each* headmate within their body feels to themselves like *his or her or their own self,* and not like the selves of the others. It would be exceedingly difficult even to express this claim without helping myself to a term like "headmate" and to a grammar that suggests that there really *are* multiple beings each enjoying a first-personal phenomenology. (Try to express it otherwise. "The plural claims that he feels like himself sometimes but not at other times" mischaracterizes the plural's claim: at every moment, they may say, each headmate feels like themselves, so there is, at any point in time, always *someone* who *does* feel like themselves. "The plural claims that he feels like multiple persons or like different people at different times" also mischaracterizes the plural's claim: the speaking headmate may say, "No—I always feel like one person, and always the same person!") But how does headmate H1 *know* that headmate H2 feels like himself—like H2—at every moment? Suppose that at a given moment, a two-headmate system is, say, cooking dinner, and headmate H1 does not feel as though it were *H1* doing it. So, headmate H1 assumes, headmate H2 must feel as though it were him (H2) doing it. Apparently H1 does not experience this very feeling—so H1 has no first-person, introspective knowledge of it. (Radden [1998], has an interesting paper on puzzles posed by simultaneous awareness of multiple consciousnesses.) The most H1 can know from her *own* experience is that certain thoughts, actions, etc., *don't* feel, to her, like *hers.* How does she know that they *do* simultaneously feel to H2's like *his?* Well, H2 could *tell* H1 that they do. But for H1 to take this "on H2's authority," H1 must *already* think H2 exists; otherwise, "H2's saying,"

---

15. I also don't know of an account of this phenomenology that is adequate to explain the sort of phenomenology that plurals describe. As Pickard (2010) points out, it's unclear that Frith's (1992) motoric account of experiences of disownership works for *thoughts* as well as it does for *actions.* Frith's account is otherwise appealing since it does not turn on the *kind* of thing someone is thinking, feeling, or doing. Pickard's account of disownership experiences is not "content-neutral" in the same way: rather, what explains someone's sense that they weren't the agent of a thought is its radical inconsistency with what she takes to be her own values, beliefs, etc. While this may be right for people with schizophrenia (the targets of both Frith's and Pickard's models), it doesn't apply to all plurals, who may say that their headmates have a lot in common, often think alike, and so on.

in inner speech, "My consciousness always feels like my own," is just another dissociated thought *of H1* occurring in her mind.

Note that this is basically just the problem of other minds. Plurals might point out that no one actually worries whether other human beings have minds, so why should one headmate worry about whether other headmates have minds? A skeptical singlet might respond that at least other human beings have perceptibly distinct bodies. But perhaps a plural would respond by asking why that should matter.

I tried to express this confusion to a few of the plurals I spoke to and the clearest answer I got was something like a "golden rule" principle: the headmate in question said, basically, that while it was true that he couldn't experience his other headmates' perspectives and thus couldn't *know* that they had perspectives in the same confident way that he knew he had his own, he would hate if his headmates started doubting that *he* had a perspective, and so wanted to extend to them the same trust he demanded from them. But once again this argument seems already to simply take as given the existence of other headmates (in this case headmates who could deny one's own sentience).

There are thus stark limits to my own understanding of the plural identity claim and accordingly to what I can make clear about that identity here. The best I can do is to try to make a little more intuitive the kind of *basis* upon which a plural comes to conclude that there is some psychological being other than the person she takes herself to be, sharing her body or mind. I will attempt this in Section 6.

The final judgment anyone should make about the truth or falsity of the plural self-belief is not pursued here. On the one hand, as I've noted in this section, there are certain difficulties in understanding the belief. Moreover, and unsurprisingly, accepting (some version of) the claim would raise a multitude of more and less obvious ethical, social, and legal difficulties. In the next two subsections, however, I explain why I think the matter does at least deserve exploration, rather than being able to be immediately dismissed.

5.2 The Metaphor Objection

The first immediate objection to engaging in any serious investigation of the truth of the plural identity claim is an objection to interpreting the plural identity claim *literally*. The objection is that people can be mistaken about the meaning of their identity claims, and that plurals might mean—perhaps, if they are rational, must mean—the plural identity claim only metaphorically, without realizing it.

The plurals I spoke to were willing to concede that the claim could be called metaphorical in one obvious sense: multiple persons within a system do not have their own bodies. As one headmate put it to me, "the *most* literal meaning of two different people would be

two completely separate people who are housed separately basically"—that is, who are not co-embodied. Granting that, however, that headmate insisted that their own plural identity claim was "about as literal as it can be with us still being in the same body."

Moreover, plurals insist that the plural identity claim is not just a metaphorical way of talking about what it is like to be a *singlet*. That is, they explicitly distinguish the claim from familiar metaphors used to talk about ambivalence, complexity, interpersonal influence, or change (e.g., *I'm of two minds about this; I'm a completely different person at work; what I said before—that was my father speaking; I don't identify with who I was then*):

> The [popular and mistaken] reasoning goes that we all have different sides to ourselves, and some people build a delusion around this natural state and come to believe that the different sides of themselves are different identities, including assigning names to these states. It's completely true that people express different sides of themselves according to different contexts. However, this is different from multiplicity. Members of a multiple group will individually experience themselves as having these "different sides," just like everyone else. (One Fox Faraday 2015, original emphasis)

One might press the metaphor interpretation by pointing to *otherkin:* people (headmates and singlets) who identify as non-human animals or fantastical creatures. Needless to say, there is something *obviously* self-contradictory in such identity claims; I take it that foxes, for instance, do not identify as foxes; certainly a fox cannot *say,* "I am a fox." Even in identifying as a fox, then, an otherkin seems to contradict that very identity claim—and this must be obvious even to them, which surely provides reason to interpret the claim non-literally. (Philosopher Katrina Haaksma has investigated the phenomenon of otherkin and suggested to me that the major meaning of the otherkin identity claim is simply that one identifies as a member of a particular human social group—the group of otherkin—with the specific creature identified with being less significant than is membership in this community, and perhaps selected on the basis of admiration or affection. Haaksma also suggested that otherkin might not *know* that this is the true meaning of their otherkin identity claims.)[16]

---

16. One question the phenomenon of otherkin poses is what it means to respect identities or other core beliefs with which one does not agree. This sort of question has a substantial philosophical aspect—involving both ethics and epistemology—especially in cases in which one thinks the beliefs may be mistaken. When they are very unusual, and especially if they bear any association to some psychopathology, the question takes on a clinical aspect. Unsurprisingly, the issue of respecting beliefs with which one does not agree is much discussed in psychiatry (e.g., Koenig [2008], on the mismatch between psychiatrists' and patients' *religious* beliefs) and in medicine more broadly (often under the umbrella of respecting cultural differences). This issue is also one that has been written about in the specific case of plural identities (see e.g., Rivera [1997:

There is overlap between otherkin and plural populations, but even if there weren't, one might take the mere fact of otherkin to provide reason against taking the plural identity claim seriously. The reasoning might be something like this: even if there were reason to *prima facie* trust or "respect" people's sincerely meant identity claims, otherkin show that some such claims are somehow metaphorical (or else obviously false). So why not take *this* strange identity claim to be one of the metaphorical (or else false) ones?[17]

The non-otherkin plurals I spoke to suggested that the otherkin identity claim was meant in some metaphorical or "spiritual" sense and explicitly distinguished it, in this way, from the plural identity claim. Some of the otherkin headmates I spoke to agreed, though one otherkin plural I raised the issue with insisted that it was literally true that one of their headmates was a fox, albeit, they acknowledged, a fox that lacked a fox's body or brain. (This exchange supported the suggestion that people may be confused about the literal meaning of the term "literal.")

It is not prima facie unreasonable to distinguish the plural identity claim from the otherkin identity claim; certainly it is possible to make two surprising claims and to mean one of them literally and the other non-literally, or for one of them to be true and the other false. But it is nonetheless the case—and the non-otherkin plurals I spoke to recognized and regretted this—that the existence of human beings who say not just "I am a headmate" but also "I am a fox" or "I am a dragon," makes it more difficult to take the plural identity claim seriously. If nothing else, otherkin identity claims create precedent for *not* accepting as literally true some identity claims that nonetheless seem to be both meaningful to and sincerely meant by those making them.

At the same time, the plural identity claim is unlike the otherkin identity claim in one crucial sense: the *literal* meaning of the claim that X is a fox is clear, forcing a metaphorical reading of a human being's claim that they are a fox. The literal meaning of the claim that X is a *person* is, however, contestable: philosophers still debate just who meets the criteria

---

33]), which some clinicians worry are not only false but also harmful or limiting in some way. Simultaneously, the voices of plurals arguing that plural identities are *not* harmful and *not* limiting—that these identities are just the *truth* of who they are—must also be listened to.

17. I also met a number of headmates who described themselves as "fictives", identifying with characters in works of fiction. This turned out to be less metaphysically interesting than I had originally thought. The fictives I spoke to said that their *being,* say, Huck Finns, meant, first, that their character and personality was based on that of Huck Finn as described in the book *Huckleberry Finn* (Twain 1884) and also that their own "life narratives" included his (or most of his) past as described in that book. Again the latter seemed more metaphysically interesting than it turned out to be, because they mostly just said that having Huck Finn's past as part of their life narrative obviously did not mean that their *body* was actually in Missouri in the first half of the 19th Century, but rather meant something more like that features of their personality and emotional life, now, were best explained *by reference to* the life story of the fictional character with whom they identified. And this itself is arguably not mysterious: identifying as someone who wasn't loved as a child, for instance, can be expected to have an effect on someone's personality, even if they *were* loved as a child. In any event, the points I am trying to make on behalf of the metaphorical objection can be made more clearly using otherkin.

for personhood and what those criteria are. And there are accounts of persons or selves according to which they are ideal, abstract, or purely intentional objects, which makes the distinction between literal and metaphorical uses of the term "person" inherently unclear.

Again, if part of what is it for it to be *literally* true that X and Y are distinct people is that X and Y are not co-embodied, then of course the plural identity claim is metaphorical. But then the question just becomes what it is a metaphor *for*, and it could turn out that, for instance, ethical questions about how to relate to systems turn more on the metaphorical meaning than on the literal one.

5.3 The Delusion Objection

The next objection to taking the plural identity claim seriously is difficult to express in a respectful manner, so I will just state it frankly. The objection is that the plural identity claim is just obviously crazy—delusional, let us say. Note that this is distinct from the objection that the plural identity claim is *false* (a claim the paper neither defends nor rejects).

One way of developing the objection might be to focus on the phenomenological abnormalities that help motivate and ground plural identity. One might simply call these hallucinations and hallmarks of psychosis; one of the articles cited earlier (Ribolsi et al. 2022), for instance, uses the language of hallucination and psychosis in speaking of the abnormal experiences and beliefs that occur with higher prevalence in people with autism. Beliefs formed on the basis of hallucination are unjustified; therefore, the plural identity claim itself is clearly unjustified.

Again this paper does not seek to engage with the truth or falsity of the plural identity claim. Indeed, given the limited ambitions of this paper, I cannot engage fully even with the delusion objection. But let me make just this one point: the basic problem with its reasoning is that the abnormal experiences that a plural takes to justify their plural identity might also be taken to ground the *truth* of that identity.

Hearing, inside one's mind, what one takes to be the voice of the King of France does not justify beliefs about the King of France nor, of course, make it the case that France has a king. But it could be *argued* that hearing, inside one's mind, the voice of a headmate justifies beliefs about that headmate and even makes it the case that there is such a headmate. Compare: hearing *my* voice in inner speech justifies some beliefs about myself and may also (help) make it the case that such a self exists in the first place. Of course, it might be questioned whether hearing my voice in inner speech justifies believing that I exist; after all, it might be said, that's a question-begging way of describing things; really I hear *a* voice in inner speech; does that justify believing that *I,* the *speaker,* exist?

Similarly, it might be questioned whether my inner speech does contribute to the *fact* of my existence; maybe even without the capacity of speech, I would have a self. But now we are getting into familiar territory about the nature of the self and its relationship to first-person narration (see e.g., Dennett, 1992, 1989). To the extent that the truth of the plural identity claim hinges on controversies in this familiar territory, it can't be dismissed as *delusional*.

It is instructive to compare the abnormal experiences of plurals with descriptions of thought insertion in schizophrenia. In discussions of thought insertion in schizophrenia, individuals with schizophrenia are described not only as judging that certain thoughts they're aware of are not their own but also as attributing them to other human beings (or other non-co-embodied agents, e.g., God). Here are two commonly cited examples of judgments of thought insertion in schizophrenia:

> I look out the window and I think that the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his. … He treats my mind like a screen and flashes his thoughts into it like you flash a picture. (Mellor 1970, 17)

> Thoughts are put into my mind like "Kill God". It is just like my mind working, but it isn't. They come from this chap, Chris. They are his thoughts. (Frith 1992, 66)

Needless to say, we do not bother to consider whether these claims could be correct. One doesn't need to know anything about Eammon Andrews or Chris (assuming both even exist) to know that they cannot put their thoughts into other people's brains somehow.

But the implausibility of plurals' claims about having other people's thoughts in their minds is more difficult to judge. There is no posited *cross-brain* thought projection being posited by a plural; the posited other person's brain just *is* their brain, and so there is no mystery about why they can experience each other's thoughts. Of course, one's immediate impulse might still be to think that it's crazy to suppose that such another person could be there, inside their same brain yet with their own thoughts. But why? The plural says something like: after all, you don't think it's crazy to suppose that *I* am here, in my brain, with my own thoughts. And if you (understandably) retreat quickly to animalism—"But you are just a human animal, singular, with one brain"—then they can say something like: but you don't deny that my brain can produce a personality, an autobiography, a center of consciousness and self-consciousness, a collection of thought patterns, preferences, and so on—and indeed, don't you admit that if I had none of these things, you wouldn't even consider me to be a person at all? So how do you know that my brain can't produce two or three or ten of these things?

Of course, it could well be that the brain *can't* do this. (Although the existence of authors who construct a number of different and richly sketched characters suggest that at least *some* brains can.) But this, in any event, is an empirical question, rather than something that can be dismissed out of hand as impossible. Or it may be that no matter how many autobiographical narratives and so on a brain produces, these must all belong to one person. But to say so is just to insist upon a class of account of the metaphysics of persons that is, notably, also rejected by some philosophers.

Indeed, the existence of philosophical accounts of personhood that allow for the possibility of multiple persons in one body recommends a certain humility in evaluating *plurals'* claims about their personhood. Dennett for instance once wrote that the idea of multiple persons or selves in one body:

> strikes many people as too outlandish and metaphysically bizarre to believe — a "paranormal" phenomenon to discard along with ESP, close encounters of the third kind and witches on broomsticks. I suspect that some of these people have made a simple arithmetical mistake: they have failed to notice that two or three or seventeen selves per body is really no more metaphysically extravagant than one self per body. One is bad enough! (Dennett 1989, 169)

In holding beliefs inconsistent with metaphysical accounts of persons that exclude the possibility of multiple personhood, both plurals and Dennett may of course be *wrong*. But surely no belief can qualify as *delusional* simply in virtue of its inconsistency with a metaphysics of persons that is controversial even amongst singlet philosophers.

6 The Origins of Plural Identity

No one is born identifying as either a singlet or a plural, but in the present cultural context, coming to identify as the sole person present in one's body is — to put it mildly — the strong default. How is it that some human beings come to identify as systems?

It seems to occur in two stages, although these likely overlap temporally in many or most cases. In the *experiential stage,* an individual has experiences *as of* there being other psychological beings inside their body or brain, whether or not they conceptualize these experiences as experiences of distinct psychological "others", much less of distinct full persons. In the *conceptual stage,* an individual comes to believe that it is possible and even fruitful to conceptualize experiences of multiplicity in terms of the existence of multiple persons sharing one brain and applies (or becomes disposed to apply) this conceptualization to their own experiences.

The order in which these two stages initiate is switched between tulpagenic and other systems. Tulpagenic plurals pass through the conceptual stage first: they read accounts from other tulpamancers, they wonder whether one truly *can* create other sentient and intelligent being inside oneself, and they find and follow instructions as to how to *generate* and *recognize* experiences of multiplicity, via the practices of tulpamancy. Often things go no further than this, since it is possible to fail to generate persuasive or systematic experiences of multiplicity. But even if one succeeds in generating such experiences, the conceptual stage may not be over. First, it's always possible to adopt deflationary interpretations of such experiences; second, even if the tulpamancer comes to view themselves as having been in some sense successful, there are successful tulpamancers who don't view their tulpas as on par with they themselves, as actual persons. Other tulpamancers however find persuasive psychological, social, and ethical accounts of tulpamancy that do urge recognition of tulpa personhood. And once the tulpamancer begins reaching the experiential stage, the two stages begin to interact: the tulpamancer attempts to apply a new conceptual framework to their experiences, and that application changes their experiences in ways consistent with that framework, thereby providing confirmation of that framework, spurring the tulpamancer on to greater engagement with tulpamancy practices, producing stronger and more compelling experiences of multiplicity.

This two-stage process and tulpamancy practices themselves are strikingly similar to the processes and practices by which many evangelical Christians in the United States come to "hear God," as documented in Lurhmann's (2012) book, *When God Talks Back: Understanding the American Evangelical Relationship with God*. One tulpamancer I spoke with remarked that "the parallels were not lost" on them, and suggested that indeed, one plausible interpretation of an evangelical who deliberately practices talking to Jesus, imagining being in Jesus' presence, attentively listening for "Jesus's voice," learning to distinguish this voice from their own thoughts, and so on, is that what they are doing is creating a *Jesus tulpa*. This Jesus tulpa might indeed be real sentient being, a real person, this headmate told me, but of course is not (unlike the actual Jesus) a real Second Temple period Jew; rather, he would simply be a person *modeled after* Jesus, with some of his characteristics (or at least with some of the characteristics that evangelical believed Jesus had).

In traumagenic (and, more broadly, non-intentionally created) systems, in contrast, experiences of multiplicity come spontaneously and first. Many non-tulpagenic plurals describe years of either not realizing how unusual were their experiences of, say, inner voices, or years of realizing that they were unusual but just understanding those experiences to be the product of, say, unusual creativity. (One "original" headmate told me that for years they thought their headmates were just fictional characters that their "brain" was coming up with, and indeed, they wrote many long works of fiction about these charac-

ters; only much later did they realize, they said, that they weren't characters but *people*.) At some point they discover—perhaps in therapy, but frequently online—the suggestion that some human brains are "home" to multiple people instead of just one. The non-tulpagenic plurals I spoke to mostly said that once they found the conceptual framework of plural personhood, things "snapped into place," without an extended period of back-and-forth between and mutual application of experience and conceptualization, but this is not universal; in fact, even within a single system, plural concepts and identity can make more sense to some headmates than to others.

The process by which a plural first comes to accept the plural identity claim, then, differs between subpopulations. But it might still seem incredible that anyone should accept such a claim. I am therefore going to describe three hypothetical scenarios that will hopefully give readers some understanding of the route by which someone might indeed come to seriously consider the plural identity claim for themselves.

The first scenario is simple. Suppose you found yourself inside a large, locked building. At a certain point, you started hearing a voice—not yours. Or you kept finding notes—not written by you. You would infer that there must be another person in the building with you. Naturally, you would assume that they were embodied, that they had their *own* body (not yours).

In the second scenario, you are still in a locked building, but this time, you begin to hear this other voice *speaking out of your mouth*.

You might be thinking that you would, in that case, reject the hypothesis that it wasn't in fact your own voice you were hearing. I assume that I would, too. But there are limits to this. Suppose that I were in the building as a subject in a neuroprosthetics experiment. Maybe one that I suspected wasn't entirely legal. Maybe that makes it more likely that someone else was using some sort of device, experimentally implanted in my brain, to speak out of my mouth, as it were.

Now this is still extraordinarily unlikely; it's the sort of conclusion that, if I accepted it too readily, would qualify me as delusional. But note that that's in part because the unlikely conclusion, in this case, is that some other *human being*—with their own body and brain—was speaking from my mouth. I cannot say just how unlikely I think that possibility is (much depends upon whether the technology that would allow it actually existed), but certainly it is exceedingly unlikely.

What plurals believe is different, though, and I find it harder to assign it a prior probability to it at all. How likely is it that some other person, *who stands in just the same relation to my brain and body that I do,* is speaking out of my mouth? That is, that there isn't some other human being, with his own vocal apparatus, who is using mine to speak —but rather someone else who has the very same claim to my apparatus that I do?

My immediate impulse is to say that it could not be true—but if you asked me why I believe that I'm the "only one home" in my brain, I would immediately appeal to evidence: if there were someone else in my brain, wouldn't they have been doing lots of things, with my body, that haven't been done? Wouldn't they be saying things—with my mouth —that surprised me and felt foreign to me? But the sort of evidence that I think I *lack* for there being another person sharing my brain is exactly the sort of evidence that plurals think they *possess*.

Now consider a third hypothetical scenario. Recently, you have been having experiences as of hearing someone else's words come out of your mouth, or as of feeling someone else act through your body, and these are causing you a lot of anxiety. In fact, you've scheduled an appointment with a psychiatrist, though you're freaked out enough to contemplate canceling. Trying to work through your anxieties the night before your appointment, you take out your journal and write, "I keep hearing words come out of my mouth that don't feel like mine. I mean it's really literally like hearing someone else speaking, except out of my own throat. I'm scared and I don't know what's happening. Am I going crazy?" The next day you bravely keep your appointment. But as you walk through the office doorway, you suddenly feel as though you are just floating within your body, rather than fully inhabiting it. You can feel your body walk to the couch and sit down, but you don't feel as though *you're* the one walking or sitting—you feel as though you're just *in* your body, with someone else moving it—sort of like a marionette, except with the strings on the inside. You hear this other voice you've been hearing—the one that doesn't feel like yours, but which issues from your mouth—introduce yourself using the childhood nickname ("Wendy") that you haven't used for years (you go by Olivia). And you hear yourself *provide an example of the concerns for which you are seeking therapy* by saying, "Last night for example, I saw myself writing this journal entry—but I swear, it was like watching someone else write it with my hand." And then as you find yourself looking down at the page in question, you hear the unfamiliar voice issuing from your mouth as it reads the entry: "Here it is. 'I keep hearing words come out of my mouth that don't feel like mine. I mean it's really literally like someone else speaking, except out of my own throat. I'm scared and I don't know what's happening. Am I going crazy?'" Then the voice from your mouth says, "I remember seeing myself write this, but it was like it was my hand writing it instead of me. And I *was* feeling afraid, except it was like it was someone else's fear. Even the handwriting doesn't look like mine. Mine is different."

You are surprised to hear yourself saying these things, which are so untrue: you remember clearly that at the moment when you were writing in your journal— *unlike* at this moment—you *didn't* feel alienated from your actions. You felt fully connected to what you were doing. And, by the way—the handwriting most certainly *is* yours!

Imagine having a number of experiences like this, until, one day, it occurs to you to wonder: in the same way that *you* think that this is *your* body, and are struggling to understand all of the words issuing from this mouth as *yours* and *yours alone*—could there also be another person, who thinks your body is *her* body, and who is trying to understand all the words issuing from your mouth as *hers* and *hers* alone?

You might have no particular view of how this is possible. You don't know how there could be someone else sharing your brain. But then—not being a neuroscientist or a philosopher—it's not as though you have a worked-out view of how *you* are "in" your brain, either.

The thought may first occur to you only as a fleeting fancy—not even a hypothesis —long before you *believe* it. But perhaps one day—feeling a bit silly—you take out a piece of paper, and write in big letters, "Wendy, are you there?" And then it seems to you that you are merely watching someone write with your hand but in foreign handwriting, "Who are you?!" And you—scarcely believing it—respond, "I'm Olivia," though it might take a number of back-and-forths before you stop feeling self-conscious, before it becomes totally natural for you to think of yourself as writing to another person. Or maybe you don't even bother to write, "Wendy, are you there?" Maybe you just *ask* it in inner speech —and then you hear—also in inner speech—"Who are you?!" Not only does it feel to you as though you were not the agent of this question, but its content startles you: to the extent that you had anticipated "hearing back" from someone else, you had expected them to introduce themselves—not to interrogate *you*.

Arguably there is no possible evidence or perhaps (though this is less certain) even any pragmatic considerations that would rationally *require* you to accept that there really *was* another person, who calls herself Wendy, inside of you. After all, there is always an alternative hypothesis: that *you* are saying and writing and thinking it all and that the fact that things seem otherwise to you merely shows that there is something *wrong* with you. Your experiences might simply be very misleading and abnormal; perhaps psychotherapy will modify them, or at least allow you to live with them while still maintaining that Wendy is just *you* in another mode. Again, this alternative hypothesis will always be available to you, and if you asked your friends, family, or therapist, it is what essentially all of them would say was the case. But this other possible interpretation—the plural premise—might occur to you too.

The story I've given above is more descriptive of DID than of tulpagenic systems, though it's not meant to offer a faithful characterization of DID multiplicity either. (Among other things, the story implies that it is only after however many decades of ordinary singlet experience that you suddenly start undergoing experiences of multiplicity.) But I think the story captures something of the phenomenological basis of plural identity just in general; as Dell (2006b) says, the phenomenology of multiplicity is in large part the

phenomenology of *intrusion*. (Although the word "intrusion" has a negative connotation that is not necessarily accurate to the experiences of multiples who are plurals.)

It might be said that it is much more puzzling how plurals could arrive at the plural premise in cases in which different headmates claim to, e.g., hear each other's voices in inner speech, since this—unlike reading written notes—is something that different persons (it might be said) cannot do, and surely plurals *know* this. But, again, headmates think that the reason that they can hear each other's thoughts is that they share a brain, as different persons ordinarily do not. Were different persons to partially share a brain (as in the case of conjoined twins Krista and Tatiana Hogan), it would not be incredible (though certainly fascinating and philosophically important) if one person could "hear" a sentence in inner speech produced by the other. Indeed, it must be said that the fact that one person cannot move another's body just by forming an intention to move or cause another person to experience a thought just by thinking something themselves is probably just a temporary technological limitation (see Jiang et al. [2019]). Yet it also seems at least a live possibility that someone whose brain and thus whose actions and mental life were subject to another person's influence in this way might still be able to distinguish (fallibly) between "self-caused" versus "other-caused" experiences and actions. (It's an empirical question, of course.)

I think we can see how this plural premise could be more appealing than the alternative *dissociated singlet premise*, especially if the former ended up somehow affording a greater degree of control over and a greater capacity to comprehend one's mental life than would the dissociated singlet premise. Suppose that, once "you and Wendy" began speaking, you managed to reach a "mutually agreed upon" policy regarding who should speak when. I'm using scare quotes here because this is, again, a question-begging way of describing the process of making this policy; if we wanted to speak consistently with a "one person per body" rule, we could say simply that, after thinking about it, you committed yourself to making certain changes in your personal speech practices (rather than saying that you and another party negotiated and came to a mutual agreement about your joint practices going forward). But again, what if you actually found it easier to make this change than you would have if you had, instead, at every moment, insisted upon your being one dissociated person—and initiated a course of intensive psychotherapy to deal with your ongoing feelings of alienation from your own experiences and actions? What if you tried psychotherapy but it didn't seem to work? Or what if the therapy itself were a further source of alienation from your experiences? Or what if your therapist didn't seem to believe or even understand what you were describing? What if it was discouraging to think of yourself in the terms your therapist used—as a fractured person

—and empowering to think of yourself as, psychologically, *just like a singlet*—a unified and unitary psychological being—albeit one whose *embodiment* was different?[18]

I don't hope to have fully illuminated the phenomenology of multiplicity, which remains on some level opaque to me still. This is partly just a general fact about phenomenology: what we can imagine is limited by the basic machinery of our own experiences (Nagel 1971). One singlet who wrote a popular media article on plurals wrote that at some point she:

> …. sat outside at a coffee shop the other day and tried to retreat within my own mind…. I wanted to see if I could imagine others milling around inside my body, but instead, I was overwhelmed by a sense of single occupancy. It felt like my mind, my selfhood, was occupying every square inch of my frame, pressing against the inside of my skull and furling out to the tips of my fingers. I waited for a voice to step out from the shadows and say hello, but there was no room for anyone else. (Telfer 2015)

This is the powerful sense I have as well. But multiples powerfully sense otherwise. My own singlet phenomenology surely does causally contribute to my belief that I am one person; mightn't it help justify that belief as well? *If* so, then perhaps the phenomenology of multiplicity can help justify plural identity.

Of course, the scenarios I described left out an important factor causally contributing to plural identity, in at least many cases: learning about other plurals. Since the "one person per body" rule is hegemonic in our culture, for many plurals (including the ones I spoke to), learning about others who rejected this rule was a necessary step in the development of their own plural identities. (Some plurals however do seem to identify as systems spontaneously; indeed, the plural who is believed to have come up with the *term* "plural" to first describe themselves, as a system, supposedly did so in the 1980s, without having encountered other systems, to their knowledge, and without having heard of multiple personality disorder.[19]) The conceptualization is like a hypothesis or theory, and experiences of feeling as though one were just watching one's body do something, experiences of hearing something in inner speech that is surprising and unrelated to what

---

18. There is also the fact that psychotherapy is time consuming and expensive. (When it is even available; it can be hard to access psychiatric care for even much more common and well understood conditions.) One traumagenic plural that I spoke to said, effectively, that, sure, maybe 20 years of psychotherapy can manage to integrate all your headmates and allow you to function effectively in daily life—or you could just come talk to *him*—a headmate in a traumagenic but functional plural system—and he could give you advice that would enable you to improve communication between and implement fairer practices among you and your headmates *this week*.

19. My source for this is personal communication, dated January 9, 2023, with the Rings System, who met and conversed with this other system (who wishes to remain anonymous) on February 2, 2019, at the Healing Together Conference in Orlando Florida.

you were just thinking about—these are the data that both support and are explained by the theory.

One could look at this contribution of sociocultural learning in different ways. When Christensen (2022) refers to "sociogenic" cases of plurality, for instance, she means to refer only to *non-traumagenic* systems, and views those cases of plurality as equivalent to (so-called) TikTok-induced functional movement disorder, a condition *distinct* from Tourette's (see e.g., Müller-Vahl et al. [2021]). This depicts "learned" plural identity as unreal some-how. But there are important sociocultural influences on experiences and expressions of voice hearing even in something like schizophrenia, which many take to be the most ob-viously biologically-based (and in that sense "real") of all the best known mental illnesses. Luhrmann et al. refer to "social kindling" as the "implicit and explicit ways in which a local social world gives significance and meaning to sensation (such as hallucination) [and that] will alter not only the way those sensations are interpreted but the likelihood and qual-ity of the sensation itself" (2015, 13). Sociocultural influences might lead to heard voices and other phenomenological abnormalities being conceived of as the voices of others—might even contribute to and alter such experiences—while those abnormalities still had independent existence.

And, crucially, it should be possible to acknowledge sociocultural contributions to plural identity even while arguing that such identities are *true*. There are sociocultural contributions to singlet identity, after all; anyway, people learn true theories from each other as well as false ones. Moreover, when we are talking about persons—who arguably owe their existence to social practices and attitudes just in general—it doesn't seem *a priori* impossible that new practices and attitudes could create new persons—headmates as opposed to human beings.

7 Conclusion

This paper has attempted to introduce a small and surprising population to the philosoph-ical community: the population of human beings who, in a sense I have tried to explain, identify as multiple people sharing a brain. I have tried to describe the contours and basis of the plural identity claim, though I have also noted some difficulties in making sense of that claim from a singlet's perspective. I have tried to distinguish the population of plurals from the population of human beings with DID and have offered some potential explanations for the overlap between plural, transgender, and autistic identities.

I have not defended the claim that a plural is multiple people or even that it is *possible* for a single human being to be multiple people, but I have argued that the claim cannot be dismissed out of hand. I have also described some scenarios meant to evoke, for sin-

glets, something of the phenomenological basis for the development of plural identity. The purpose of doing this has been to try to illuminate plurals' own perspectives on their mental lives and identities. An investigation of the actual metaphysical status of the plural identity claim would have to tackle not only metaphysical debates about the nature of persons generally, as well as psychological and (perhaps) neural facts about plurals, but also the deeply ethical nature of our concerns about personal identity, since almost all of our ethical assumptions and attitudes presuppose the "one person per body" rule that plurals claim does not apply in their cases.

References

American Psychiatric Association. 2022. *Diagnostic and Statistical Manual of Mental Disorders*. Dsm-5-tr. American Psychiatric Association Publishing. https://doi.org/10.1176/appi.books.9780890425787

Amorpha Household. n.d. "Faq." http://dreamshore.net/amorpha/faq.html

Arnaud, Sarah. 2020. "Self-Consciousness in Autism: A Third-Person Perspective on the Self." *Mind & Language* 37 (3): 356–72. https://doi.org/10.1111/mila.12356

Batman, Lucia (FreyasSpirit), and Knapp (Irenes) Irene. n.d. "The Plurality Playbook." https://www.pluralpride.com/playbook

BenitoFlakes_. 2021. "Comments on the 2018 Census." The Tulpas Subreddit on Reddit. https://www.reddit.com/r/Tulpas/comments/oa287n/comments_on_the_2018_census/

The Blackbirds. n.d. "The Layman's Guide to Multiplicity." http://www.karitas.net/blackbirds/layman/whatlike.html

Braude, Stephen E. 1995. *First Person Plural: Multiple Personality and the Philosophy of Mind*. Rev. ed. Lanham, Md: Rowman & Littlefield Publishers.

Butler, Lisa D. 2011. "Must Dissociation Be Unusual?" *Journal of Trauma & Dissociation* 12 (4): 454–56. https://doi.org/10.1080/15299732.2011.570234

Callum, James. 2019. "Answer Log." https://www.quora.com/Do-your-other-alters-ever-say-think-or-do-things-that-are-the-type-of-thing-you-yourself-would-say-think-or-do/answer/James-Callum-4/log

Chevallier, Coralie, Gregor Kohls, Vanessa Troiani, Edward S. Brodkin, and Robert T. Schultz. 2012. "The Social Motivation Theory of Autism." *Trends in Cognitive Sciences* 16 (4): 231–39. https://doi.org/10.1016/j.tics.2012.02.007

Christensen, Emily M. 2022. "The Online Community: DID and Plurality." *European Journal of Trauma & Dissociation* 6 (2): 100257–58. https://doi.org/10.1016/j.ejtd.2021.100257

Dell, Paul F. 2006b. "A New Model of Dissociative Identity Disorder." *Psychiatric Clinics of North America* 29 (1): 1–26. https://doi.org/10.1016/j.psc.2005.10.013

Dell, Paul F. 2006a. "The Multidimensional Inventory of Dissociation (MID): A Comprehensive Measure of Pathological Dissociation." *Journal of Trauma & Dissociation* 7 (2): 77–106. https://doi.org/10.1300/J229v07n02_06

Dell, Paul F. 2011. "An Excellent Definition of Structural Dissociation and a Dogmatic Rejection of All Other Models." *Journal of Trauma & Dissociation* 12 (4): 461–64. https://doi.org/10.1080/15299732.2011.570236

Dennett, Daniel C. 1989. "The Origins of Selves." *Cogito* 3 (3): 163–73.

Dennett, Daniel C, and Nicholas Humphrey. 1989. "Speaking for Ourselves: An Assessment of Multiple Personality Disorder." *Raritan* 9 (1): 68–98.

Dennett, Daniel C. 1992. "The Self as a Center of Narrative Gravity." In *Self and Consciousness*, edited by Frank S. Kessel, Pamela M. Cole, Dale L. Johnson, and Milton D. Hakel, 103–15. Psychology Press. https://doi.org/10.4324/9781315807652

Flynn, Em. 2011. "Plurality for Skeptics." 2011. https://www.exunoplures.org/main/articles/medicalisation/skeptics/

Fried, Ina. 2019. "Longtime Google Engineer Known for Speaking Out About Sexual Harassment and Controversial Projects Is Leaving." Axios. https://www.businessinsider.com/google-engineer-fong-jones-leaves-company-2019-1

Frith, Christopher Donald. 1992. *The Cognitive Neuropsychology of Schizophrenia*. London: Taylor, Francis. https://doi.org/10.4324/9781315785011

Garrett, Aiden/Estelle. 2023. "Living, Together: A Singlet's Guide to Plural Allyship." https://sexualhealthalliance.com/nymphomedia-blog/living-together-a-singlets-guide-to-plural-allyship

Gleaves, David H. 1996. "The Sociocognitive Model of Dissociative Identity Disorder: A Reexamination of the Evidence." *Psychological Bulletin* 120 (1): 42–59. https://doi.org/10.1037/0033-2909.120.1.42

Hardcastle, Valerie Gray, Owen Flanagan, and The Hegeler Institute. 1999. "Multiplex Vs. Multiple Selves: Distinguishing Dissociative Disorders." Edited by Sherwood J. B. Sugden. *Monist* 82 (4): 645–57. https://doi.org/10.5840/monist199982431

Huntjens, Rafaële J. C., Madelon L. Peters, Liesbeth Woertman, Loes M. Bovenschen, Roy C. Martin, and Albert Postma. 2006. "Inter-Identity Amnesia in Dissociative Identity Disorder: A Simulated Memory Impairment?" *Psychological Medicine* 36 (6): 857–63. https://doi.org/10.1017/s0033291706007100

Jaswal, Vikram K., and Nameera Akhtar. 2019. "Being Versus Appearing Socially Uninterested: Challenging Assumptions About Social Motivation in Autism." *Behavioral and Brain Sciences* 42: e82. https://doi.org/10.1017/s0140525x18001826

Jiang, Linxing, Andrea Stocco, Darby M. Losey, Justin A. Abernethy, Chantel S. Prat, and Rajesh P. N. Rao. 2019. "Brainnet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration between Brains." *Scientific Reports* 9 (1): 6115–16. https://doi.org/10.1038/s41598-019-41895-7

Kindt, Merel, and Marcel Van Den Hout. 2003. "Dissociation and Memory Fragmentation: Experimental Effects on Meta-Memory but Not on Actual Memory Performance." *Behaviour Research and Therapy* 41 (2): 167–78. https://doi.org/10.1016/s0005-7967(01)00135-8

Kluft, Richard P. 1991. "Clinical Presentations of Multiple Personality Disorder." *Psychiatric Clinics of North America* 14 (3): 605–29. https://doi.org/10.1016/s0193-953x(18)30291-0

Kluft, Richard P. 2006. "Dealing with Alters: A Pragmatic Clinical Perspective." *Psychiatric Clinics of North America* 29 (1): 281–304. https://doi.org/10.1016/j.psc.2005.10.010

Koenig, Harold G. 2008. "Religion and Mental Health: What Should Psychiatrists Do?" *Psychiatric Bulletin* 32 (6): 201–3. https://doi.org/10.1192/pb.bp.108.019430

Laursen, Christopher. 2019. "Plurality Through Imagination: The Emergence of Online Tulpa Communities in the Making of New Identities." In *Believing in Bits*, 163–79. Oxford University Press. https://doi.org/10.1093/oso/9780190949983.003.0010

Luhrmann, T. M. 2012. *When God Talks Back: Understanding the American Evangelical Relationship with God*. 1st ed. New York: Alfred A. Knopf.

Luhrmann, Tanya M., R. Padmavati, Hema Tharoor, and Akwasi Osei. 2015. "Hearing Voices in Different Cultures: A Social Kindling Hypothesis." *Topics in Cognitive Science* 7 (4): 646–63. https://doi.org/10.1111/tops.12158

Marsh, Rosemary J., Martin J. Dorahy, Chandele Butler, Warwick Middleton, Peter J. De Jong, Simon Kemp, and Rafaele Huntjens. 2021. "Inter-Identity Amnesia for Neutral Episodic Self-Referential and Autobiographical Memory in Dissociative Identity Disorder: An Assessment of Recall and Recognition." Edited by Vedat Sar. *Plos One* 16 (2): e245849. https://doi.org/10.1371/journal.pone.0245849

Martin, Jean-Remy, and Elisabeth Pacherie. 2013. "Out of Nowhere: Thought Insertion, Ownership and Context-Integration." *Consciousness and Cognition* 22 (1): 111–22. https://doi.org/10.1016/j.concog.2012.11.012

Mellor, C. S. 1970. "First Rank Symptoms of Schizophrenia: I. The Frequency in Schizophrenics on Admission to Hospital II. Differences between Individual First Rank Symptoms." *British Journal of Psychiatry* 117 (536): 15–23. https://doi.org/10.1192/s0007125000192116

Michal, Matthias, Heide Glaesmer, Rüdiger Zwerenz, Achim Knebel, Jörg Wiltink, Elmar Brähler, and Manfred E. Beutel. 2011. "Base Rates for Depersonalization According

to the 2-Item Version of the Cambridge Depersonalization Scale (CDS-2) and Its Associations with Depression/anxiety in the General Population." *Journal of Affective Disorders* 128 (1–2): 106–11. https://doi.org/10.1016/j.jad.2010.06.033

Mikles, Natasha L., and Joseph P. Laycock. 2015. "Tracking the Tulpa." *Nova Religio* 19 (1): 87–97. https://doi.org/10.1525/nr.2015.19.1.87

Müller-Vahl, Kirsten R, Anna Pisarenko, Ewgeni Jakubovski, and Carolin Fremer. 2022. "Stop That! It's Not Tourette's but a New Type of Mass Sociogenic Illness." *Brain* 145 (2): 476–80. https://doi.org/10.1093/brain/awab316

Nagel, Thomas. 1971. "Brain Bisection and the Unity of Consciousness." *Synthese* 22 (3/4): 396–413. https://www.jstor.org/stable/20114764

Nijenhuis, Ellert R. S., and Onno Van Der Hart. 2011. "Dissociation in Trauma: A New Definition and Comparison with Previous Formulations." *Journal of Trauma & Dissociation* 12 (4): 416–45. https://doi.org/10.1080/15299732.2011.570592

Olson, Eric T. 2014. "The Metaphysical Implications of Conjoined Twinning." *The Southern Journal of Philosophy* 52 (S1): 24–40. https://doi.org/10.1111/sjp.12073

One Fox Faraday. 2015. "Open Letter." June 2015. https://www.tumblr.com/1ff/122106386443/open-letter

Pickard, Hanna. 2010. "Schizophrenia and the Epistemology of Self-Knowledge." *European Journal of Analytic Philosophy* 6 (1): 55–74. https://hrcak.srce.hr/63456

Radden, J. 1998. "Pathologically Divided Minds, Synchronic Unity and Models of Self." *Journal of Consciousness Studies* 5 (5–6): 658–72. https://www.ingentaconnect.com/content/imp/jcs/1998/00000005/f0020005/901

Radden, Jennifer. 1996. *Divided Minds and Successive Selves: Ethical Issues in Disorders of Identity and Personality*. Cambridge, MA: MIT Press

Reuben, Katherine E., Christopher M. Stanzione, and Jenny L. Singleton. 2021. "Interpersonal Trauma and Posttraumatic Stress in Autistic Adults." *Autism in Adulthood* 3 (3): 247–56. https://doi.org/10.1089/aut.2020.0073

Ribolsi, Michele, Federico Fiori Nastro, Martina Pelle, Caterina Medici, Silvia Sacchetto, Giulia Lisi, Assia Riccioni, Martina Siracusano, Luigi Mazzone, and Giorgio Di Lorenzo. 2022. "Recognizing Psychosis in Autism Spectrum Disorder." *Frontiers in Psychiatry* 13 (February): 768586–87. https://doi.org/10.3389/fpsyt.2022.768586

Ribáry, Gergő, László Lajtai, Zsolt Demetrovics, and Aniko Maraz. 2017. "Multiplicity: An Explorative Interview Study on Personal Experiences of People with Multiple Selves." *Frontiers in Psychology* 8 (June): 938–39. https://doi.org/10.3389/fpsyg.2017.00938

Riesman, Abraham. 2019. "The Best Cartoonist You've Never Read Is Eight Different People." https://www.vulture.com/2019/03/lb-lee-dissociative-identity-disorder-comics.html

Rivera, Margo. 1997. *More Alike Than Different: Treating Severely Dissociative Trauma Survivors*. University of Toronto Press. https://doi.org/10.3138/9781442664845

RSpacefox. 2021. "The Solution to Medicalists in the Plural Community." https://twitter.com/RSpacefox/status/1471133024330960898

Schiltz, Hillary K, Alana J McVey, Bridget Dolan Wozniak, Angela D Haendel, Rachel Stanley, Alexis Arias, Nakia Gordon, and Amy Vaughan Van Hecke. 2021. "The Role of Loneliness as a Mediator between Autism Features and Mental Health among Autistic Young Adults." *Autism* 25 (2): 545–55. https://doi.org/10.1177/1362361320967789

Shinyuu. 2015. "Reddit Subreddit Tulpas." The Tulpas Subreddit on Reddit. https://www.reddit.com/r/Tulpas/wiki/census//#wiki\%5F2015

Sinnott-Armstrong, Walter, and Stephen Behnke. 2000. "Responsibility in Cases of Multiple Personality Disorder." *Philosophical Perspectives* 14: 301–23. https://www.jstor.org/stable/2676134

Stronghold. 2021. "Plurality Vs Plural Experiences." https://powertotheplurals.com/plurality-vs-plural-experiences/

Taylor, Marjorie, Sara D. Hodges, and Adèle Kohányi. 2003. "The Illusion of Independent Agency: Do Adult Fiction Writers Experience Their Characters as Having Minds of Their Own?" *Imagination, Cognition and Personality* 22 (4): 361–80. https://doi.org/10.2190/ftg3-q9t0-7u26-5q5x

Telfer, Tori. 2015. "Are Multiple Personalities Always a Disorder?" https://www.vice.com/en/article/vdxgw9/when-multiple-personalities-are-not-a-disorder-400

Thompson, Nathan. 2014. "The Internet's Newest Subculture Is All About Creating Imaginary Friends." https://www.vice.com/en/article/exmqzz/tulpamancy-internet-subculture-892

Turell, Susan C., Christopher Wolf-Gould, Sana Flynn, Silver Mckie, Matthew A. Adan, and The Redwoods. 2023. "It's Just a Body: A Community-Based Participatory Exploration of the Experiences and Health Care Needs for Transgender Plural People." *European Journal of Trauma & Dissociation* 7 (4): 100354–55. https://doi.org/10.1016/j.ejtd.2023.100354

Twain, Mark. 1884. *The Adventures of Huckleberry Finn*. Various Publishers.

Tye, Michael. 2003. *Consciousness and Persons: Unity and Identity*. Representation and Mind. Cambridge, Mass: MIT Press

US Surgeon General. 2023. *Our Epidemic of Loneliness and Isolation*. https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf

Vessière, Samuel. 2016. "The Hypnotic Nature of Human Sociality, Personhood, And Interphenomenality." In *Hypnosis and Meditation: Towards an Integrative Science of Conscious Planes*, edited by Amir Raz and Michael Lifshitz, First Edition. Ox-

ford, United Kingdom: Oxford University Press. https://global.oup.com/academic/product/hypnosis-and-meditation-9780198759102

Williams, David. 2010. "Theory of Own Mind in Autism: Evidence of a Specific Deficit in Self-Awareness?" *Autism* 14 (5): 474–94. https://doi.org/10.1177/1362361310366314

Zarzycka, Beata, and Dariusz Krok. 2021. "Disclosure to God as a Mediator between Private Prayer and Psychological Well-Being in a Christian Sample." *Journal of Religion and Health* 60 (2): 1083–95. https://doi.org/10.1007/s10943-020-01107-3

Žikić, Olivera, Snežana Ćirić, and Marija Mitković. 2009. "Depressive Phenomenology in Regard to Depersonalization Level." *Psychiatria Danubina* 21 (3): 320–26. https://hrcak.srce.hr/49216

cognethic.org