A Thesis Presented to
The Faculty of Alfred University

Dismissing and Salvaging the Metaphysics of People

by

Conrad Cheung

In Partial Fulfillment of
the Requirements for
The Alfred University Honors Program

April 30, 2018

Under the Supervision of:

Chair:
Dr. Emrys Westacott

Committee Members:
Professor Linda Sikora
Dr. Hope Childers
Dr. Robert Stein

ABSTRACT

Disputes over personal identity — over the metaphysics of people — are old wars. Philosophers feud over the classic questions: *What makes something count as a person? What does it take for someone at different times to count as the same person?* The assumption that runs through these disputes is, simply put, that there are facts of the matter waiting to be discovered and described by the right theory of personal identity.

Every part of this assumption, I argue, is mistaken. Questions about personal identity are *not* answerable as a 'matter of fact'; they're answerable only as questions about our linguistic and conceptual conventions. What's to blame is epistemic inaccessibility: we have no good grounds for believing that there's any right theory of personal identity. We can't, therefore, say what the metaphysical facts about people are, or indeed whether there are any at all. Our most sensible recourse, I suggest, is to adopt a *pragmatic* view: to describe personhood or persistence as a property of any entity x is nothing more than to say that the relevant predicate — "is a person", "is the same person" — is contextually assertible of x. We should accept this view on grounds that are themselves pragmatic: with no justification to make any firmer metaphysical claims, we should regard the shifty, stipulative, ordinary concept of "person" as *good enough*. It can still do the everyday work we want it to.

ACKNOWLEDGMENTS & A BITE-SIZED OVERVIEW

OK — here goes. I probe here a set of very old questions about the metaphysics of people, or about *personal identity* — questions that are at least three centuries old, since Locke (1690/1998) published on the matter, and that are still under dispute. The result of the probing is a wholesale dismissal of the longstanding philosophical enterprise, on the grounds that it faces fatal problems of epistemology: we have no way, it turns out, of knowing what the metaphysical facts about people are, or whether there are any at all. The rundown follows.

The questions at issue are easily stated and easily motivated. First, the problem of *personhood*: what makes something count as a person? More carefully: what is the state of affairs φ such that for any x, x is a person just if φ? Second, the problem of *persistence over time*: what does it take to count as the same person over time? Or: what is the required state of affairs σ, such that for any x and any y, x at t_1 is the same person as y at t_2 just if σ?

Two views, each having survived the gauntlet of historical disputes and each purporting to give the right answers to the classic questions, dominate the contemporary field. One is the Psychological View, which says roughly that personhood consists in and persists as overlapping chains of psychological continuity. In a crude slogan: we go where our minds go. The other is the Biological View, which says that personhood is nothing over and above having a human body that's continuously alive. The slogan: we go where our bodies go. Variants of each view abound, but all fall neatly into either of the two broad camps. The two sides argue endlessly, relying characteristically on two kinds of moves: appeals to intuition and appeals to theoretical

virtues. But it can look, both to outsiders and to insiders, that there's been hardly any progress in decades: neither side has made any serious concessions since the first contemporary articulations of their views. And when we notice, for instance, that epistemic communities in the natural sciences typically converge on theories at a far faster rate than do philosophers (three whole centuries for personal identity and no dice!), the philosophical disputes can start to seem a little ... awry. We are, after all, just talking about what people are. All the while, biologists have, by consensus, a fairly precise definition of what *humans* are. What's going on?

I start by zooming out from the disputes and asking whether there might be a higher-order explanation for the philosophers' deadlock. Four diagnoses are suggested by the literature; I name them the *just-more-work*, the *nihilist*, the *semantic*, and the *epistemic* diagnoses. The just-more-work diagnosis is optimism about the metaphysical project; make no mistake, it says, there *are* facts of the matter about personal identity, and with *just more work*, we'll find them eventually. The nihilist diagnosis is grimmer. It starts with the independent premise that there are, in fact, no such things as people. (Trust me that the view, fully explained, isn't as preposterous as it at first sounds.) The thought then goes: if it's true that there are no people, then there are no entities that claims of personal identity are *about*; and so the deadlock can be chalked up to the disputants' failure to recognize that their claims don't refer to anything. But the last two candidate diagnoses are where things really get interesting.

The semantic diagnosis has it that the disputes are "merely verbal" — that if we look hard, we'll find that the disagreements over personhood and persistence are over nothing more than terminological choices, and not over what there really is. Consider an example: if you're the sort of person who indiscriminately calls any drink in a V-shaped glass a martini, and I, the snobbier between us, classify as martinis only those drinks with a particular ratio of vodka, vermouth, and olive juice, then we'll appear to have a substantive metaphysical disagreement if there should happen to be a V-shaped glass of apple liqueur in front of us: you'll say that there's a martini, and I'll say there

isn't. And thanks to our different definitional schemes, we'll appear to have spun a deep disagreement about the existence of a martini. I admit openly that this diagnosis is attractive, on the surface. But, as I show, it loses plausibility once scrutinized.

The last diagnosis — the epistemic one — is the one I favor. In the course of building the case for it, we will see that if it's right, none of the other diagnoses can be right. The epistemic diagnosis maintains that the real trouble afflicting the disputes is that we have no way to tell which theory of personal identity is correct. So what explains the philosophical deadlock is *undecidability* about the facts of personal identity: we have little to no justification for thinking that personhood or persistence is one way or another — i.e., is what the Psychological View says it is or what the Biological View says it is.

My argument is roughly procedural. I canvass all the ways we can tell whether a theory is right, or likelier than its competitors to be right: by its *empirical adequacy*, by its *accordance with intuition*, and by its *theoretical virtues.* These are the standard *theory-choice criteria.* (Occasionally, ordinary-language inference is counted as a criterion, but I quickly get it out of the way by giving a brief demonstration of its nonviability as a guide to truth.)

Armed with our theory-choice criteria, I examine the dispute between defenders of the Psychological and Biological Views. What's contested between the two camps is *what the metaphysical facts about people are.* What we want to know is whether, by applying our theory-choice criteria, we can determine whether one view or the other is likelier to be right. The answer turns out, after process of elimination, to be *no.* And rather worryingly, the problems we encounter with each theory-choice criterion are not easily resolved; they are deep, structural difficulties. I draw here heavily on the work Rudolf Carnap (1950) and his modern-day intellectual heirs to pinpoint what's gone wrong. A harrowing implication follows from the failure of the theory-choice criteria: since we have no way to tell which of the views is right, we have no way to tell what the metaphysical facts about people are.

4

I then turn to a nearby debate: that between the *realist* about people and the *nihilist* about people. What's contested between the two is *whether there are people.* I rerun the procedure used on the previous debate. The result is the same: considerations of empirical adequacy, accordance with intuition, and theoretical virtues all fail to tell us whether one side or the other is likelier to be right. This time, the consequence is even worse. If we can't tell whether people exist, we can't tell whether there's anything that claims of personal identity are *about.* And if that's right, then, by inference, we can't tell whether there are metaphysical facts about people. This second failure amplifies the previous failure: we're not just unable to say *what* the facts are; we're unable to say *whether* there are metaphysical facts about people at all.

This twofold failure is, I think, no accident. I detail why I think so, before turning to the most salient objection to my case: that the trouble is not really epistemic, but semantic. This objection advances the semantic diagnosis, described earlier, as an explanatorily superior alternative to my epistemic diagnosis. I reject the semantic diagnosis, after finding that the notion of the "merely verbal" dispute — *the* crucial notion that the semantic diagnostician relies on — doesn't, when formulated precisely, apply to the debates over personal identity.

A positive account of personal identity and a reinterpretation of the longstanding disputes conclude the paper. In the wake of the serious troubles that beset the standard theories, I say that we should adopt a *pragmatic* view of personal identity, for reasons that are themselves pragmatic. The account I formulate aims to track our ordinary concept of "person", with the result that there's nothing metaphysically essential or fixed about personhood or persistence. Roughly, the pragmatic view amounts to the following. On personhood: what it takes for something x to be a person is just, in the context at hand, for the linguistic predicate "is a person" to be assertible of x — and nothing more. And on persistence: for x and y to count as the same person is just, in the context at hand, for the predicate "is the same person" to be assertible of x and y — and nothing more. Put more casually: personhood and persistence are the stuff of

semantics, not metaphysics — nothing more than say-so attributions made sensible and actionable by our ordinary sociolinguistic behavior. (This should not, I have to point out, be confused with the semantic diagnosis. Our objects of analysis aren't the same: the semantic diagnostician is talking about the nature of the *disputes* over personal identity; I am talking about the nature of *personal identity*.)

Now, if all that's right, what are we to make of the (evidently irresolvable) disputes? I say that we would do well to reinterpret the philosophical enterprise as *normative*, rather than descriptive — as talk about how personhood and persistence *ought* to be defined, not about what they are. Doing so is more charitable than the alternative — trashing the enterprise altogether — and gives us a soberer perspective on what's really at stake in our ivory-tower discussions — among other things, the granting of legal status, the shape of public opinion, and the content of social norms.

TABLE OF CONTENTS

1. The big fuss over personal identity

1.1. Reconstructing the first-order theories

A nice and easy way to come across classic metaphysical problems is to ask what vindicates some of our most boring, workaday presuppositions. With enough prodding, the classic questions of personal identity — questions about the metaphysical nature of people — eventually turn up this way. (Or so, at least, I like to think.) Here's the issue of personhood, so conceived: you and I are people, but oranges and underwear are surely not. What — if anything — makes this so? And here's the issue of persistence over time: since you started reading this paragraph, your metabolism has taken you through plenty of microphysical changes. But it's plain old common sense that you are still, now, the person who started reading this paragraph seconds ago. What — if anything — ensures that sameness of person?

These two questions take simple general forms:

PERSONHOOD: What makes something a person?

PERSISTENCE: What does it take for someone at different times to count as the same person?

Occasionally, the philosopher of personal identity finds herself concerned with a third question —

EXISTENCE: Are there people?

— but those embroiled in the usual disputes over PERSONHOOD and PERSISTENCE, we will see, take for granted that the answer to the EXISTENCE question is obviously *yes*.

We will, for now, grant that assumption and shelve the EXISTENCE question. (Not to worry: it will make an important reappearance in §3.3.)

To discern more clearly what's at issue, let us begin with a semi-formal revamp of the PERSONHOOD and PERSISTENCE questions. Here, first, is the bare form of a personhood claim:

For any x, x <u>is a person</u> just if ...

The question we want to ask is:

PERSONHOOD*: What is φ, such that for any x, x is a person just if φ?

Here, second, is the bare form of a persistence claim:

For any x and any y, x at time t_1 <u>is the same person as</u> y at time t_2 just if ...

In question form:

PERSISTENCE*: What is σ, such that for any x and any y, x at t_1 is the same person as y at t_2 just if σ?[1]

φ and σ are *satisfaction conditions* for personhood and persistence — states of affairs under which it's true, respectively, that x is a person or that x and y are the same person. More carefully: when we ask the question "what is φ?" we are asking what

---

[1] Here are the full formalizations:

PERSONHOOD**: $\forall x\, (Px \Leftarrow \varphi)$, where P = 'is a person'

PERSISTENCE**: $\forall x \forall y\, (Rxy \Leftarrow \sigma)$, where R = 'is the same person as'

the world must be like for the relevant predicate — 'is a person' — to be considered true of, or satisfied by, the relevant entity — any entity that x ranges over. And likewise for σ: what must the world be like for the two-place predicate 'is the same person as' to be considered true of, or satisfied by, x at t_1 and y at t_2?[2]

Now schematized, the questions of PERSONHOOD* and PERSISTENCE* need answers. Here's a quick sketch of the contemporary views on offer — the surviving candidates from over three centuries of philosophical dispute. There are two majority views.[3] On one side is the Psychological View, which maintains roughly that personhood consists in and persists as overlapping chains of psychological continuity; on the other side is the Biological View, which claims roughly that personhood is nothing over and above having a human body that is continuously alive.[4] (I'll henceforth abbreviate the Psychological View as PV and the Biological View as BV.) Both views have variants: some revise PV to include extra requirements for personhood, like causal links between mental events (Shoemaker 1991); some make BV put the burden of persistence on the brain or parts of the brain, rather than on the whole human organism (Puccetti 1973; Campbell & McMahan 2010).

Despite their obvious difference, PV and BV share an underlying formal structure: they're both examples of *essentialism* about personal identity. Let's expose precisely what this consists in, to get a clearer view of where the two views converge and diverge. Begin by recalling the question with which we began §1.1. We're people, but oranges and underwear aren't; what makes that the case? There's one way to go about an answer which seems natural and obvious (but is, importantly, not necessarily right). It goes like this: in order to vindicate the intuition that all (or most) of us humans are people and most other kinds of entities aren't, start with the idea that certain

---

[2] Just for completeness, here's EXISTENCE*: Does there exist at least one x such that x is a person?

[3] See Bourget and Chalmers (2014) for the data.

[4] See Parfit (1984) for what I take to be the contemporary paradigm of the Psychological View, and Olson (1997a, 2003) for a clear and consistent representation of the Biological View.

characteristics exclusive to humans[5] *make that the case.* Again, a little differently: start

by presupposing that one or some of the features that we humans have in common but

that other kinds of entities lack are *essential* to personhood. Depending on the preferred

view, those features might include "a bundle or collection of different perceptions"

(Hume 1739/2000), rational agency (Rovane 1998), self-awareness (Locke 1690/1998),

a living body (Olson 1997a, 2003), or a properly functional brain (Campbell &

McMahan 2010). And once the philosopher makes the essentialist assumption — *that*

there exist features essential for personhood — her task is just to sleuth out *what* those

features are.

But what exactly does it mean for certain features to be *essential* for personhood?

For a given set of features $\{f_1 ..., f_n\}$ to be essential to some entity x's personhood is

for it to be necessarily the case that if x has $\{f_1, ..., f_n\}$, then x counts as a person. This

if-then relation goes by the name "grounding" in the literature, and gets phrased in

many ways: e.g., the fact that x is a person *is grounded in* the fact that x has the

relevant features; x's having those features *makes it the case* that x is a person; x's

being a person *depends on* x's having those features; x is a person *in virtue of* x's having

those features; x's having those features *is a criterion for* x's being a person; and so

on. Which formulation is used doesn't matter; they all denote the same type of

dependence relation.[6] What's more, the particular metaphysics of modality involved in

essentiality — here the modal operator "necessarily" — can be cashed out in many

ways. For instance, following Lewis's (1968) canonical proposal, we can say that the

set of features $\{f_1, ..., f_n\}$ is essential to x's personhood if and only if it's the case that

at every possible world w_n, if x has $\{f_1, ..., f_n\}$ in w_n, then x counts as a person. But

---

[5] And possibly also, of course, to species or members of species similar to humans.

[6] It's typically accepted that grounding is a one-way relation that is *irreflexive* (nothing grounds itself)*, transitive* (if P grounds Q, and Q grounds R, then P grounds R), and *asymmetric* (if P grounds Q, it's not the case that Q grounds P). On a standard view, P grounds Q, or $Q \Leftarrow P$, if and only if Q is less fundamental than, and metaphysically depends on, P. There's been no shortage of literature that tries to precisify this; see, e.g., Rosen (2010), Bennett (2011), Schaffer (2009), and Berker (forthcoming).

we need not commit ourselves to the Lewisian program to make sense of essentiality; I will remain neutral here about which semantics is most appropriate to institute.[7]

We can now construct an essentialist template for personhood claims:

ESSENTIALISM ABOUT PERSONHOOD: Necessarily, x is a person if and only if Fx, where F is the predicate 'has the set of features $\{f_1, …, f_n\}$'.

Much of the heavy lifting here is done by the 'necessarily' operator, which captures the core essentialist idea that there are certain properties that an entity *must* have to count as a person, or alternately, that it cannot lose without ceasing to be a person. All the work left now is to see how PV and BV fill out the template. Roughly, what the defender of PV maintains is that $F = $ 'has the relevant psychological features'. The resulting view thus runs:

THE PSYCHOLOGICAL VIEW ON PERSONHOOD: Necessarily, x is a person if and only if x has the relevant psychological features.

Exactly *which* psychological features count will depend on the substantive details of PV's particular formulation. In his paradigm articulation of PV, for instance, Derek Parfit (1984: 207) specifies as his criteria "overlapping chains of strong psychological connectedness, itself consisting in significant numbers of direct psychological connections like memories, intentions, beliefs/goals/desires, and similarity of character".

---

[7] However, this is *not* to say that any semantic program will work. For the standard essentialist approach to come out intact (i.e., *qua* essentialism as it's normally understood), the chosen semantics should be tethered to one form or another of metaphysical realism. What plausibly *won't* capture the spirit of essentialism, as it's meant here, is an analysis which cashes out essentiality in exotic, anti-realist ways (see, e.g., Ásta 2013).

Contra the advocate of PV, the defender of BV maintains that $F =$ 'has a properly functioning human body'. So we get:

> THE BIOLOGICAL VIEW ON PERSONHOOD: Necessarily, x is a person if and only if x has a properly functioning human body.

On the matter of personhood, the dispute between defenders of PV and BV centers *solely* on this question: what features must x have — what features will x have essentially — if x is to count as a person?

Let us run through the same procedure for the matter of persistence.

> ESSENTIALISM ABOUT PERSISTENCE: Necessarily, x at t_1 is the same person as y at t_2 if and only if Rxy, where R is the two-place predicate 'shares the set of relational features $\{r_1, ..., r_n\}$ with'.

The defender of PV says that $R =$ 'has the relevant kinds of psychological continuity with', such that

> THE PSYCHOLOGICAL VIEW ON PERSISTENCE: Necessarily, x at t_1 is the same person as y at t_2 if and only if x has the relevant kinds of psychological continuity with y.

And the defender of BV claims that $R =$ 'has the same human body as';

> THE BIOLOGICAL VIEW ON PERSISTENCE: Necessarily, x at t_1 is the same person as y at t_2 if and only if x has the same human body as y.

So far, so mechanical the legwork. Now, how do the actual disputes between defenders of PV and BV play out?

1.2. What's the fuss like?

Defenders of PV and BV tend to follow patterned lines of argument, relying characteristically on two kinds of moves: appeals to intuition and appeals to theoretical virtues. At times the disputants aim to show that their view cleaves closest to folk intuition, or that the opposing view yields conclusions that violate it. So goes a classic thought experiment from Locke (1690/1998), in which he defends an early version of PV by imagining that a prince and a cobbler have swapped memories. Locke intuits that the prince and the cobbler would 'migrate' with their memories: each would consider himself to *be* the person to whom the memories had originally belonged, rather than the person whose body they now occupied.[8] Thus he concludes that the criterion for persistence over time is one's collection of memories. In a rough thesis: necessarily, x at t_1 is the same person as y at t_2 if and only if x has the same memories as y.

Likewise pumping intuitions, Olson (1997b) offers a challenge from the opposing side, arguing that PV forces a counterintuitive answer to the question in his paper's title: "Was I Ever a Fetus?" The answer from the advocate of PV, he maintains, must be *no*. Olson's line of thinking runs like so: current embryology tells us that an early-stage human fetus has virtually no psychology; and since adult people therefore bear no psychological continuity with their fetus selves, the advocate of PV must conclude that "[n]o person was ever a fetus, and no fetus ever comes to be a person."[9] And such

---

[8] See Williams (1970) for a counterexample. His thought experiment will make a reappearance in §3.2.

[9] Here's the argument, more slowly. On the standard form of the Psychological View, for someone to count as the same person over time, she must maintain psychological continuity. According to current embryology, an early-stage human fetus has virtually no psychology. We can't be psychological continuous with an entity that lacks a psychology, so we're not psychologically continuous with our early-stage fetus selves. So we must conclude, Olson says, that "[n]o person was ever a fetus, and no fetus ever comes to be

a conclusion strikes most as intuitively preposterous, so the PV defender must either concede that her view is mistaken, or else "bite the bullet" and aver that the consequences of her view, while true, are utterly bizarre.

At other times, disputants charge the opposing view with lacking theoretical virtues, such as parsimony and elegance. The tellingly named "Problem of Too Many Thinkers" is one such charge. Some versions of PV, the Problem of Too Many Thinkers points out, are committed to a distinction between a person and a human organism. This is because according to PV, the persistence conditions for people and human organisms are different: while someone x may lack personhood because they lack the relevant psychological features, x may nonetheless be a human organism — as in the case of a fetus. And once x gains the relevant psychological features, x becomes a person. But a strange ontological commitment is embedded in this distinction between people and humans: if PV is right that people and humans are distinct kinds of entities, then most of us — you and I — are each an overlap of two distinct entities — a person *and* a human organism — that somehow manage to spatiotemporally coincide, despite their observational indistinguishability.[10] This weird consequence of *colocation* flouts not just common sense, but the principle of parsimony: it multiplies entities beyond necessity.[11]

1.3. Why does the fuss persist?

For terminological flexibility, I will call PV and BV the *standard accounts* of personal identity, and the ruckus raised between their defenders the *standard disputes.*

---

a person." And such a counterintuitive conclusion should motivate believers in the Psychological View to abandon it, or at least get seriously worried about its tenability.

[10] Compare the "Problem of Too Many Thinkers" (Hershenov 2006, 2013) with Olson's "Indiscernibility Problem" (2001), which poses a similar basic threat. Both problems ask: how can entities with different persistence conditions — say, a person and a human organism — be composed by the exact same particles at the exact same spatiotemporal coordinates?

[11] Strange issues of causal overdetermination are also involved here; see, e.g., Merricks (1998).

The standard disputes can look deeply frustrating: why is it so hard to settle the matter of what makes us people? Biologists, by consensus, have a set of criteria for something to count as *human*; genetic markers establish membership in the primate group *Homo sapiens.* What keeps metaphysicians from achieving the same for *person*?

Here are four *prima facie* plausible diagnoses — a roundup of those that have been offered, or more often taken for granted, in the literature. First, an optimistic possibility: the problem is merely a lack of the relevant knowledge. Maybe a yet undiscovered neurological process is what grounds personal identity, and once it's discovered, metaphysicians will quickly come to a consensus. If that's the case, we're simply waiting to discover the right criteria for personhood and persistence. All we need is more work. Call this happy idea the *just-more-work* diagnosis.[12]

Then again, our problem might be of an altogether different kind. It could be linguistic: maybe metaphysicians are being duped by some fiendishly subtle ambiguity in the questions they argue over, or more trickily, their answers to the questions of personal identity merely *seem* incompatible because they're formulated in the languages of different conceptual schemes. In the latter case, a difference in the disputants' terminological choices causes unspoken disagreement over *what the right expression* for some state of affairs is, but not genuine disagreement over what the state of affairs itself is.[13] The resulting dispute is "merely verbal", like

> the dispute between the purist who says that only cocktails made of gin or
> vodka, dry vermouth, and perhaps an olive or two count as martinis, and the
> sorority girl who calls practically anything a martini as long as it served in the
> classic V-shaped glass. If these two are seated at a table on which such a glass

---

[12] This is a bit awkward, I admit; it's a play on Putnam's "just-more-theory manoeuver". I had originally called this the *realist* diagnosis, but that is misleading, since, for instance, the semantic diagnosis can accommodate realism about personal identity with no trouble.

[13] The jargon for this phenomenon is *quantifier variance*. See Hirsch (2002) for an analysis.

contains some nonsense made of sour green apple liqueur, the latter will say

that there is a martini there, and the former will deny it. (Bennett 2009: 13)

Notice how different the tack of this diagnosis is from the tack of the just-more-work diagnosis. It maintains that nothing substantive is really at issue in the dispute between defenders of *PV* and *BV*; all that constitutes the disagreement is how each side chooses to talk about *the exact same referent*.[14] If this is in fact the kind of problem the disputants are having, there could very well be facts of the matter about personal identity, but we're far too caught up in linguistic quarrels to see them. Call this the *semantic* diagnosis.

There remain even more radical possibilities, of which I'll cover two. One starts with *nihilism* about people: the thesis that there are no such things as people. Philosophers such as Peter Unger (1979a,b) and Trenton Merricks (1998) have defended nihilism. Suppose, skipping over the arguments for it, that it's right. Then our ordinary talk of people would be plainly false, fictional, or meaningless; then, too, the failure to find the right account of personal identity would have an easy explanation: there is no right account *to* find, since there doesn't exist anything for an account to be right *about*. This phenomenon goes in the literature by the rather inelegant name "catastrophic presupposition failure". It is worth detailing the problem with some care. Not all statements involving a referring term — here "person" — will necessarily come out false or meaningless if "person" turns out not to have a referent that exists (i.e., *people*). Consider, for instance, the statement "Now Billy's a person you can trust!" Yablo (2006) would point out that what that sentence is *about* is not Billy's personhood, but Billy's trustworthiness; and we can motivate such an intuition by noticing that the asserted meaning of the statement is preserved if we substitute "a person" with another term whose referent's existence is not being challenged, like

---

[14] Eli Hirsch (2002, 2005) gives what is probably the most sensitive articulation of this kind of diagnosis. See also Chalmers (2011) and Bennett (2009).

"someone" or "a dude" or "a human being". But not all statements keep their meanings intact when a term fails to refer; indeed, all statements that are *about* personal identity will plausibly turn out false or meaningless, if the nihilist is right that people don't exist. Such is the case with a simple statement of personhood: "Billy's a person." Personhood is what the sentence is *about*; one cannot make a substitution for "person" without losing the statement's asserted meaning. It's *these* kinds of statements that turn out meaningless or false (or perhaps without a truth-value at all; see Kleene 1952/2009) if the nihilist is right. And since such statements of personhood and persistence are exactly what philosophers of personal identity take as their primary objects of analysis, the nihilist's being right would be devastasting ("catastrophic") for the philosophers' enterprise.[15] The nihilist diagnosis looks grim. If true, it would mean that metaphysicians have been engaged in purely make-believe disputes, and that no matter how hard they try, they *could not have* genuinely metaphysical disputes over people. But let's take a step back from the precipice.

Here's a last possibility: we have no way of telling whether any of the standard accounts is right, or likelier to be right. On this view, the lack of progress on personal identity can be chalked up to undecidability between competing accounts; they are, preferability-wise, on an equal footing (Williams 1970; Sider 2001). Notice that this is

---

[15] More to the point, the reason that the phenomenon is called "catastrophic presupposition failure" is that the fact that people don't exist would <u>catastrophically</u> result in the <u>failure</u> of claims about personal identity which <u>presuppose</u> that fact. I use the fact that assertions depend on presuppositions but not vice versa (see fn 26) to start my argument in §3.1; also, see Yablo (2006). You may, following Yablo, be wondering whether nihilism about people, if true, would in fact result in *non*-catastrophic presupposition failure, and so not give the outcome I describe. For instance, you might think that upon revelation that the nihilist thesis is true, the search for criteria of personal persistence could simply be reinterpreted having been as a search for criteria of persistence of something *like* or *coincident with* people all along, with no serious consequence for claims of personal identity.

This isn't right. Consider a plausible candidate for the person-substitutable object of claims about personal identity — say, *human.* If the nihilist is right, then a non-catastrophic presupposition failure would simply have us rephrase the questions of personal identity as questions of *human* identity. But by the standard accounts of personal identity, humans differ from people in their essential makeup (see end of §1.2), and thereby differ in their conditions for personhood and persistence. Plausibly, then, claims about personhood and persistence couldn't simply be rephrased to circumvent the catastrophe.

not the same as the nihilist view; it doesn't imply that *there cannot be* right answers to the questions of personhood and persistence. But it doesn't contradict nihilism either. Nor, even, does it contradict the just-more-work view that there is some fact of the matter about personhood and persistence that we could discover. What this diagnosis does, instead, is to make a sober assessment: *there could be* right answers, but we don't know what they are. And moreover, while there could be right answers, we wouldn't be justified in believing either that there *are* right answers or that there *aren't*. So it's perfectly *possible* that the just-more-work view or the nihilist view (or any of the positions nearby) is right about the metaphysical state of affairs, but we can't say (Bennett 2009). Call this the *epistemic* diagnosis. I will be saying plenty more about it.

The way the standard disputes go, it should be clear, presupposes that the *just-more-work* diagnosis is right, or at least most favorable. The nihilist diagnosis makes the project of finding grounds for personhood and persistence nonsensical, while the semantic and epistemic diagnoses offer unflattering aerial views from which it looks like the disputants have been confused by deep and persistent mistakes. Let us see what exactly the just-more-work diagnosis — I'll call it the *standard assumption*, since participants in the standard disputes take it for granted — consists in.

The standard assumption is two-tiered. In so-called "first-order" metaphysics, which is concerned with what there is in the world[16], the assumption says that there are right answers. This is to say that there exist perfectly determinate facts about what grounds personhood and persistence, and that those facts are describable in terms of sets of criteria. *Make no mistake*, says the just-more-work diagnostician, *there are answers!* And what's more, we ought to trust that it's possible to formulate those answers with enough conceptual and linguistic precision that they can be evaluated as true or false.

---

[16] At least, as W. V. O. Quine (1980) canonically put it in the essay "On What There Is". This formulation, traceable back to Plato, isn't without detractors; see Schaffer (2009) for a revamp of the dissenting Aristotelian view that metaphysics concerns *how the world is structured.*

Thus do metaphysicians get billed as seekers and finders of deep, serious, substantive truths: what the criteria for personhood and persistence *really are*, whether people *do* exist, what kind of thing the referent of "me" or "she" *is*, and so on (Unger 1979b; Shoemaker 1991; Baker 2002; Burke 2003; Olson 1997a, 2003). This first-order view is, in a word, *realism* about matters of personal identity.

In "second-order" *meta*metaphysics, which is concerned with how the first-order metaphysical debate is going, the standard assumption has it that the metaphysics of personal identity is *descriptive*. This is to say that what metaphysicians are doing when they argue about and around matters of personal identity — that is, what the whole philosophical enterprise concerning personal identity *is* — is aiming at an acceptable formulation of descriptive answers — of those first-order criteria that tell us the facts of the matter about personhood and persistence.[17] This second-order view is *descriptivism* about the philosophical debate over personal identity.

1.4. Disentangling the fuss

We've just surveyed the MO in the metaphysics of personal identity — how the standard assumption plays out in the standard disputes between the standard accounts. All our "isms" are in place: we've seen that the standard assumption is that of *realism* about personal identity; that the standard disputes take *descriptivism* for granted about the metaphysician's task; and that the standard accounts are examples of

---

[17] My intended sense of "descriptive" here shouldn't be confused with that in P. F. Strawson's influential distinction between descriptive and revisionary metaphysics. *He* meant to distinguish between those metaphysical (descriptive) accounts that take reality to be accurately reflected in our existing conceptual scheme and those (revisionary) accounts that take the world to be accurately reflected in some other, better conceptual scheme we ought to adopt. When I talk about "descriptive criteria", I mean nothing more than criteria that are meant to accurately reflect real states of affairs concerning personhood, persistence, etc., no matter whether the conceptual schemes those criteria are packaged with do or don't differ from our existing conceptual scheme.

essentialism about personhood and persistence. This neat package of views — this entire approach — is, I aim to show, a grand mistake.

The case I'll be making connects three accusations: the standard accounts can't tell us what the facts of the matter are; the standard assumption is thus misguided; and the standard disputes turn out, as a result, to be proceeding on mistaken grounds. My proposal in the wake of these accusations is that we should adopt a pragmatic view about personal identity, for reasons that are themselves pragmatic. I'll say in due course what exactly this position amounts to, but we can get the gist in a slogan: to be a person is nothing more than to participate in a useful sociolinguistic fiction about "people".

My case rests on the "epistemic diagnosis" I suggested in §1.3: we have little to no justification for thinking that any of the *existing* standard accounts of personal identity is right, nor justification for thinking that *any* standard (i.e., essentialist) account of personal identity could be right. We thus have no good grounds for regarding personhood, persistence, or any other aspect of personal identity as some fact of the matter about ourselves which we merely, as it were, aim to describe. Understood more reasonably, aspects of personal identity are *nothing more* than concepts that we've stipulated according to the purposes and ends of our ordinary practices and that, therefore, vary in meaning by context. And that's where the standard assumption goes wrong. There *is* something that vindicates the ordinary belief that we, but not pomegranates and socks, are people, but that "something" is *not* a set of essential features we have. It's rather, in a literal hearing of the cliché, *just semantics*: that "something" is the collection of context-variant rules that govern how we use the concept and the word "person".

Naturally, repercussions follow: if the standard assumption is implausible, the standard disputes proceed on an implausible assumption. Matters of personal identity are, seen soberly, *not* matters concerning what there really is. What we face, then, is the question of how to characterize what metaphysicians are actually up to when, in

21

the standard disputes, they feud endlessly about and around matters of personal identity. A charitable view I favor is that their disputes are best seen as *normative*, not descriptive — as tussles over how the concept of "person" *ought* to be used, what the grounds for personhood *ought* to be, and so on. Reconsidered in this way, the standard disputes are less like disputes in the natural sciences and more like disputes in ethics: they aim not at discerning facts about the world, but at persuading us about which facts we should prefer as the basis of our claims about people.

For reasons that we'll see, I find the alternative diagnoses — the just-more-work, the semantic, and the nihilist — unconvincing. The just-more-work and nihilist diagnoses suffer from the metametaphysical problems that afflict PV and BV (detailed in §3). And while the semantic diagnosis no doubt offers a superficially attractive explanation for what has gone awry in the standard disputes, it looks far from plausible upon close scrutiny (discussed in §4).

My case gets built up piecemeal. Here's a guide. §2 begins with some fun; it aims to put in perspective what the point of our ordinary concept of "person" is. This happens by way of a thought experiment: how nightmarish would our lives be without presuppositions about personhood and persistence propping up our everyday practices? §3 starts the argument in earnest, against the standard assumption and for the epistemic diagnosis. I canvass all the ways we might be able to determine whether an account of personal identity is true, or likelier to be true than its competitors: by ordinary language, by empirical evidence, by intuition, and by the theoretical virtues. Applied to the standard accounts, all four fail. This failure leaves us with no way of knowing which account of personal identity is right, and so no way of knowing what the metaphysical facts about people are. I then reapply the lessons learned from this first failure to a nearby debate: that between the realist and the nihilist about people. That application results in a second failure, so that we fail to know not only *what* the metaphysical facts about people are, but *whether* there are metaphysical facts about people at all. A good deal of modesty about personal identity proves appropriate;

22

advocates of the standard accounts, of the just-more-work diagnosis, and of the nihilist diagnosis turn out to be entirely unwarranted in the confidence of their positions. §4 entertains the most salient objection to the epistemic diagnosis — namely that the semantic diagnosis, on which the standard disputes are "merely verbal", has greater explanatory power. I defend the view that while matters of personal identity are themselves semantic phenomena, the reason for "all the fuss" — for the intractability of the disputes — is fundamentally epistemic, not semantic. §5 sets out my pragmatic alternative. If we can't say that there's any fact of the matter about personal identity, what should we think about our ordinary claims, intuitions, and practices that involve the concept of "person"? I extend a clue picked up in §2: propositions about personhood, persistence, and other aspects of personal identity are stipulations we make *for* various purposes — and nothing more. I accordingly offer pragmatic formulations of personhood and persistence. To wrap up, I suggest a helpful way to recast the longstanding standard disputes in a new light: what the disputants are really up to is normative work. Arguments over personal identity are, on this view, "metalinguistic negotiations" (Plunkett & Sundell 2013; Thomasson 2016), exchanges in a field of so-called "conceptual ethics" (Burgess & Plunkett 2013a,b). To think that the standard disputes aim at anything 'deeper' or more revelatory should invite serious doubt.

2. The point of personal identity, via a thought experiment

2.1. The S-World

Let's start with garden-variety beliefs about personal identity — claims that seem so obviously true that it looks absurd to challenge them, and absurder yet to deny them. We've got at least one about every key aspect of personal identity:

Existence: There are people in the world.

Personhood: We're people. Other entities, like the waffles we had for breakfast, aren't.

Grounds of personhood: Something makes it the case that we, but not waffles, are people.

Persistence: We're *also* going to have waffles for dinner! That is: we who had waffles for breakfast are the very same people who will have waffles for dinner.

And the list goes on. Normally, raising a challenge to these sorts of claims — and even just stating them explicitly — strikes us as puzzling and pointless, because the truth of them is already taken for granted in many of the ordinary things we do. The claims fit into a larger picture: they're pieces and links in the conceptual backdrop of our everyday business. If they're to be competently performed, all kinds of activity, from waxing nostalgic about childhood to applying for jobs, require that we presuppose certain things about our personal identity, such as our continuation as the same people over time.

Let's get a more vivid picture of how practically indispensable these intuitions are, by way of a thought experiment. Imagine that there's an absurd alternate world in which ordinary claims about personal identity aren't presuppositions, but openly acknowledged, controversial ideas. Whether any of them is true or false, no one is too sure. Call it the Skeptical World, or S-World for short: a world where, all else being

equal[18], we and everyone else have suddenly become hardcore skeptics on matters of personal identity. By "hardcore" I mean that our skepticism in the S-World is not just intellectual but dispositional: not only what we believe, but everything we say and do involving the concept of "person", is tempered by self-doubt and second-guessing. Any S-Worlder who happens to suspect she might have alighted on a preferable view of personal identity can never quite get confident enough to endorse it, and so goes on wondering.[19]

Faced regularly with questions about the nature of people, our S-World selves flounder. On the issue of the grounds of personhood — *in virtue of what does an entity count as a person?* — we hesitate to answer. On the matter of *our* personhood, we're bizarrely uncertain: *do there really exist the right kinds of facts to establish that we're people?* And on the issue of persistence — *what does it take for someone to count as the same person at different times?* — we waffle over candidate criteria: psychological continuity, bodily continuity, maybe persistence of a soul, maybe some other inscrutable state of affairs that guarantees identity[20] . . . *who knows?*

We S-Worlders get constant anxiety over the weirdest of metaphysical possibilities. Just as some of our real-world counterparts find plausible the idea that we're in a Matrix-like simulation (Bostrom 2003), we entertain seriously some radical claims about personal identity: maybe it's the case that we're brand new people all the time, despite the fact that we look like, feel like, think like, and have the same names, birthdates, and birthplaces[21] (among other things) as the people we used to be. (Let

---

[18] For the most part. We might need to turn skeptical about some of our *other* beliefs for the S-World scenario to plausibly obtain — on pain of rational incoherence. Adjusting our confidence levels in those other beliefs, as needed, should suffice.

[19] Stipulating changes in our basic psychology may also be necessary for this set-up of the S-World to look plausible. I trust you'll grant me whatever changes are necessary.

[20] See Zimmerman (1998) for his coverage of the "identity mystics", philosophers like Merricks (1998) who take the view that personal identity (specifically persistence) can't be explained or grounded by more fundamental facts or conditions.

[21] At least for the sake of the wacky view being described, it's plausible to say that birthdate and birthplace are facts concerning *human organisms* rather than people. So we can expect someone's birthdate

me mention that such a thought is already conditional on the idea that we *are* people — an idea which is, in the S-World, not without controversy.) How could it be that we're always brand new people? Perhaps the slightest change in someone's microphysical makeup — say, the shedding of a single carbon atom from her skin — suffices to destroy the person she is and give rise to a new person (Hume 1739/2000). The consequence would be that each of us is a human organism that materially coincides with a succession of people going in and out of existence at virtually every consecutive point in time.[22] (Some, like Sider (1996), say that we are "stages" of a person, each distinguished by the particular spatiotemporal coordinates we occupy.)

Since we S-Worlders lack any firm view and any easy intuitions about what conditions do and don't make someone count as the same person, we find it hard to stave off the worry. Some of the heavier brooders among us fret about what practical consequences follow from this possibility that we are never — or, more carefully put, are at no two differentiable points in time — the same people. For instance, shouldn't we be continually applying for new licenses, visas, and other pieces of ID if we're always losing personal identity with — if, put more loosely, we're no longer *really* — the people they were issued to?

Seeing beyond its apparent silliness, you may think the question rests on a basic mistake — one concerning qualitative and numerical identity. (For two things to be *qualitatively* identical, the philosophical distinction goes, is for them to have qualities in common. But for them to be *numerically* identical requires a strict logical or metaphysical equivalence: they have to be the exact same thing.[23] Standardly,

---

and birthplace to be fixed by the circumstances of her human birth and to be unaffected by the fact of her having undergone ("undergone" in some very loose sense) the instantaneous creation and destruction of many people.

[22] A tidier representation of this "succession", idealized on the assumption that a new person always comes into existence at every point in time, might go:

someone = { $\langle person_1,\ time_1 \rangle$, $\langle person_2,\ time_2 \rangle$ ... $\langle person_n,\ time_n \rangle$ }

[23] Here's some elaboration for the nonphilosopher, using examples. Qualitative identity is the kind of relation found between identical twins: similarity in virtue of the fact that $twin_1$ and $twin_2$ share a

philosophers have thought that personal identity over time, or persistence, is *numerical* identity.) Your objection runs something like this:

> Look, you strange S-Worlders — suppose it really *is* the case that you're constantly becoming new people. There's still an obvious pragmatic reason why you wouldn't have to apply over and over again for new papers. The identifying facts used in ID — name, gender, birthdate, facial appearance, and the rest — are able to establish qualitative identity *at best*. And barring cases of successful identity fraud, qualitative identity works reliably to match, and to detect mismatch between, the human represented on paper and the human presenting the papers. So given that qualitative identity is good enough for the practical purposes of issuing authorities — indeed, is the best identification standard they've got — why in the world would they be compelled to care about a metaphysical thing like numerical identity, over and above qualitative identity?

Unfortunately, there's a catch. The S-World's brand of mind-bending skepticism afflicts *everyone.* The issuing authorities aren't themselves without doubt about what the right standard for identification is, or ought to be. They squabble internally about how the rules and regulations, if revised to reflect the skepticism they hold in common with the legislature and the public, would look. When customs officers are tasked with checking that the traveler standing before them 'is the same as' the traveler described on paper, they wonder whether they should really be aiming for a higher standard of evidence, which takes account of the metaphysical state of affairs — proof of *numerical* identity. They recognize higher-level reasons why that worry might be warranted: maybe the very international practice of IDing overvalues practical considerations.[24]

---

proportion of their properties. But *numerical* identity is more demanding: for the twins to be numerically identical, such that $twin_1 = twin_2$, would require that they be *one and the same person.* Leibniz's Law, though not uncontroversial, formulates this as indiscernibility between entities.

[24] Of course, this may strike you as obviously misguided. Whether it does is quite beside the point. The point is that skepticism about those parts of our existing conceptual scheme that have to do with

Not just border control, but all kinds of government agencies throughout the S-World, are plagued and made inefficacious by not-so-simple worries like these. And in many places, public efforts to amend the rules for identification are already underway. Of course, no one is too sure whether numerical identity is the right standard either, but the concern at hand is the fact that the existing rules take too much for granted about the metaphysical facts: is there anything more substantial than practical expedience, asks the good S-Worlder, that would justify our believing that qualitative identity checked off a document could suffice to count as personal identity?

I'll remind us that despite the tangle of concerns so far raised, the practical question we sought to answer is simple. If it's true that we are new, numerically distinct people at every point in time, would we always need new visas and licenses? The short answer is, well, *maybe.* Many factors — unknowns about how the rest of the S-World is — preclude a decisive answer.[25] Still others keep our S-World selves from knowing how seriously we should take the possibility. One problem, the type of which we'll meet again in this paper, is the epistemic challenge of how any of us could come to know whether it's really the case that our ideas about personal identity, wacky or levelheaded alike, are true.

We've just looked at one quandary of personal identity — one out of millions or billions, plausibly, mulled over and negotiated daily in the S-World. Other cases are just as odd, or worse. Here are two more, briefly. First: in court, criminal defense lawyers capitalize on S-World judiciaries' internal doubts about their rules and norms of adjudication. Some make statements that wouldn't stand a chance in the real world:

---

personal identity, and skepticism about our prior commitments propping up those parts, is sufficiently intelligible that inhabitants of the S-World could have it.

[25] The answer depends on how realistically we model the whole of the S-World and its possible futures — i.e., how much of the psychological, social, political (etc.) facts of our real world are stipulated to hold in the S-World. In a somewhat realistic S-World, we might expect practical considerations to prevail by, say, the phasing out of "person" talk in favor of talk about entities whose facts we're more confident about, like human organisms. Or alternately, we might expect the abovementioned "public efforts" to reform existing rules about ID, or proposals to change any law with objectionable presuppositions about personal identity, to go nowhere, likewise for practical reasons.

It may be the case, Your Honor, that the overwhelming majority of evidence suggests the defendant's guilt. Yes: the DNA matches; the footage shows him at the scene of the crime; the testimonies unequivocally single him out. But we can acknowledge all this while maintaining his innocence. For the person who committed the crime is not my client: it was a person he *used to be*, but that person has since lost numerical identity with him, and no longer exists.

And here's our last case. At breakfast, it regularly occurs to Joanne that various things in her home — her dresser, her plates, the delicious waffles she's eating — might be eligible for personhood. The only trouble is that she can never quite make sense of how things in the world would have to be for that to be true.

2.2.   What's the point?

Life in the S-World is hard. Making explicit and raising a challenge to our intuitions about personal identity has, as the S-World thought experiment suggests, the effect of putting our existing practices — social, legal, political, and other — at risk of paralysis and dysfunction. Thus are ordinary claims of personal identity *practically indispensable* in the real world. That practical indispensability, I've tried to show by example, comes in virtue of two key features: ordinary claims of personal identity are (i) *presuppositional* and (ii) *purposive*. That is: (i') we *make as if* ordinary claims of personal identity are true, no matter whether in fact they are, and (ii') we do so for *practical reasons and ends*. The conjunction of (i) and (ii) accounts for why it seems preposterous for anyone to deny a basic claim about persistence like, "I'm the same person on Wednesday as I was on Tuesday," and likewise for why the possibilities entertained by S-Worlders strike us as peculiarly wrong: that we're just successions of instantaneous people and so always need new IDs; that someone who's qualitatively

identical with a criminal could nonetheless deny sameness of person and so exculpate himself; or that the inanimate objects in Joanne's home might count as people, given the right circumstances.

Seen from a higher altitude, the conjunction of (i) and (ii) works as a two-tier structure of justification. First, the *presuppositional* status of ordinary claims about personal identity makes us feel warranted in judging S-World-type claims to be implausible. This is because to make a presupposition is to stipulate a set of truth-conditions — what things in the world would have to be like for the presupposition to come out true — and fairly or unfairly, it's against these *ad hoc* truth-conditions that we judge S-World-type claims for their truth or falsity (von Fintel 2004). Second, the *purposive* status of ordinary claims about personal identity explains why their adoption as presuppositions seems warranted — viz., because they serve our everyday interests, and we wouldn't, as S-Worlders can attest, manage too well without them.

We've seen two particular reasons for stipulating standards of personhood and persistence: we want to identify the right travelers and charge the right criminals. But backing up those reasons are, arguably, broader and more fundamental reasons. One of them is sociolinguistic: we need *some* kind of shared conceptual scheme and vocabulary (i.e., those centered on the "person" concept) to handle the theoretical burden of the many things — actions, rights, responsibilities, obligations, power, etc. — that we want to attribute to other human organisms.[26,27] And this sort of attribution, to be clear, can be cashed out in purely linguistic-conceptual terms; we need not make any special metaphysical commitments to the existence of actions,

---

[26] We may also want to attribute such person-related rights, responsibilities, etc. to *non*-human entities, such as orangutans and chimpanzees; see Feltman (2014). This attests, in my view, to the flexibility of the "person" concept: it does what we want it to.

[27] One way to get a loose intuition for the sort of theoretically flexible placeholder concept that "person" is is to notice its occurrence in phrases of ordinary language — though I wouldn't want to put too much pressure on them. Take, for instance, the implied attribution of status — moral, social, and legal standing — to *people* in loaded sentences like "*Fs* are people too!" or "*Gs* aren't people", where *Fs* and *Gs* are members of recognized social units, like corporations or racial or gender populations.

rights, responsibilities, etc. to make sense of it. Plausibly, in other words, the sociolinguistic motivation for all sorts of ordinary statements about people is that we need some concept — "person" — which stipulates an entity that is able to *satisfy certain predicates,* such as 'has a moral responsibility'. Those predicates could attribute certain kinds of status, as in "x has the right to µ" or "x is obliged to ρ", or they could give relational descriptions, as in "y has more power than x".

A second reason for personal-identity talk is psychological: we need *some* way to identify and explain what patently feels for each of us, in our day-to-day or second-by-second experience, like a unitary 'source' of behaviors, sensations, thoughts, desires, and other physical and mental states (Hume 1739/2000; Carey 2011) — at least in those cases where other candidate concepts like "body", "mind", and "consciousness" aren't sufficiently generic or don't quite fit the conceptual bill. Indeed, both this psychological explanation and the sociolinguistic explanation appear to track in the ordinary concept of "person" a constitutive cognitive 'primitive': the concept of "agent".[28] An agent, says Carey (2011), is characterized by "goals, communicative interactions, attentional states, and causal potential" — just the kinds of properties whose attribution our ordinary concept of "person" works to accommodate in our everyday linguistic behavior.[29]

So far what I've done is to sketch what the point of the ordinary concept of "person", and of ordinary claims about personal identity, is. We make presuppositions about the metaphysical nature of people for practical purposes. But most philosophers engaged in the standard disputes over personal identity will likely find this characterization too weak. No doubt our ordinary conceptual scheme about "person" is *useful*, they want to say, but what about whether it's *true?* Matters of personal identity are far more than pragmatic, they insist: claims about people are, on their view, claims

---

[28] See Carey (2011) for a rundown of other primitive concepts, such as "object" and "cause".

[29] One might suggest, on the side, that a better candidate concept than "person" for the psychological reason is "self".

about a certain kind of thing bearing essential features, and accordingly, claims about personal identity are claims about what the fact of the matter regarding that certain kind of thing is. Let us see if this is so.

3. Metametaphysical trouble on the horizon

3.1. Ordinary-language inference and theory-choice criteria

To say that we *presuppose* personhood and persistence as facts about ourselves is to say that we treat them as if they were true, for the sake of something else — an assertion, a belief, a policy statement, a law. But whether they're in fact true is an altogether separate matter. To demonstrate this, let's articulate the claim more precisely: the truth or falsity of a presupposition *does not depend on* the truth or falsity of the thing — the assertion, the belief, the rule — for the sake of which the presupposition was made. Now, suppose that you and your friend are arguing over the personhood of machines, and your friend exclaims:

(R) Hey — smart robots are people too!

Among the presuppositions in her assertion (R) are:

(1) Some entities besides smart robots are people.
(2) Some robots are smart.
(3) There are robots.
(4) There are people.

Now: if the truth of a presupposition actually *did* depend on the assertion for whose sake it was made, then making (R) false should make the presuppositions (1)-(4) false too. But if we try it — say smart robots are *not* people too — we find that the truth of the presuppositions remains intact: it can still be the case that some entities besides smart robots (like humans!) are people, that some robots are smart, that there are robots, and that there are people.

Because presuppositions are true or false *independently* of the assertions, beliefs, and rules they're meant to prop up, the mere fact that we ordinarily say, believe, and do things that presuppose that we have personhood and persistence cannot tell us whether we really do. In other words, looking at ordinary language — at our everyday, pre-philosophical claims — can't tell us whether there are metaphysical facts about people, or what, if anything, those facts are.

How, then, can we know about personal identity? Surely, after all, the standard philosophical accounts of personal identity — PV and BV — purport to be telling us the facts of the matter; how could we know whether one of those accounts is right? What's salient to note here is that this is merely an instance of a general question: how can we know whether *any* theory is true, or more likely to be true than its competitors? Classically, three types of criteria have mattered (Kriegel 2013): the theory's (a) empirical adequacy, (b) accordance with intuition, and (c) theoretical virtues. Consider, for instance, historical candidates for a conception of gravity. Among them, Newton's law is likelier to be true than Aristotle's theory, given its superiority in at least two of the three respects mentioned[30]: the law of universal gravitation (a') gives a unifying description of normal (non-relativistic) empirical cases with vastly more accuracy and coverage and (c') exhibits more simplicity insofar as it postulates fewer entities (compared to Aristotle's explanation by way of the four elements, "natural places"

---

[30] I say *two* because I hesitate to say whether Newton's law accords better with intuition than Aristotle's explanation. Indeed, I think there's significant reason to doubt that intuition is an appropriate type of consideration for scientific (as opposed to metaphysical) theory choice. See more in §3.2.

towards which each element moves, "celestial spheres" where those natural places are located, etc.).

These three criteria could factor into our judgment of which theory of personal identity, and therefore which claims about the metaphysics of people, is likeliest to be right. What is at stake in such a judgment is hard to overstate, since the theory we judge to be right about any given subject matter tells us what we ordinarily think of as facts about that subject matter — as the phrase goes, "the fact of the matter". And for a claim to be a *fact* is, on a typical understanding[31], for it to correspond with the way things actually are in the world, or part of the world. Can the three theory-choice criteria help us figure out which of the standard accounts of personal identity is likeliest to be right?

3.2. Between the standard accounts: knowing what the facts are

I'll start by spilling the beans: *no.* We have a battery of strong and independent reasons to doubt the viability of each criterion.

Let us first consider *empirical adequacy.* A theory's empirical adequacy is a measure of how well it accounts for the empirical evidence — how well it can explain and predict observed phenomena.[32] It's primarily on empirical adequacy that the "success" of successful scientific theories rests: the considerable power of Newton's law to inform us about ubiquitous and common events (say, how apples fall) is its most obvious boon over the Aristotelian explanation. But while empirical adequacy in a *scientific* theory gives us good reason to think it's true, it looks doubtful that a *metaphysical* theory of personal identity could demonstrate empirical adequacy at all. To see this, suppose that defenders of the standard accounts are locked in disagreement over whether

---

[31] That is: on the *correspondence theory of truth.*

[32] Some include *retrodiction* — the ability of a theory to account for, or 'reverse-predict', past events — as a feature of empirically adequate theories; see Kriegel (2013).

Johnny, a coma patient who has just woken up with a drastic change in personality and loss of memory, is the same person as pre-coma Johnny. This is a question of persistence. The defender of *BV* answers *yes*, he's the same person, since he's retained his body. But the defender of *PV* says *no,* given Johnny's discontinuous and altered psychology. By the lights of empirical adequacy, who's likelier to be right?

Neither, it seems — for it's impossible to show that either side actually *explains* or *predicts* anything about Johnny's case. The problem can stated simply. A scientific theory explains any event *E* by saying why *E* happens; predicts *E* by telling us that, given the right circumstances, *E* will happen; and has its explanation or prediction *confirmed by observation of E itself happening.* In Johnny's case, each disputant will want to say that their pet theory can explain and predict the relevant event: the defender of *PV* will point to Johnny's psychological discontinuity as the cause of his change in person, while the defender of *BV* will point to Johnny's bodily survival as the reason for his sameness of person. But since the relevant event — change or preservation of person, as the case may be — is an unobservable metaphysical event, we end up with no observational evidence to confirm the explanations/predictions. So how do we know whether either side is right? Well, we *can't* — not by empirical means. The two standard accounts are *empirically equivalent* (Paul 2012).[33]

Next, let's turn to intuition. Using intuition as a guide to whether a theory is likely to be true has been a philosophical staple since Classical antiquity (Plato 380

---

[33] A clever but confused objection, taking after the insights in J. L. Austin's (1962) speech act theory, might be tempting as a last resort. It goes something like this: the Psychological or the Biological View can exhibit a peculiar kind of *quasi-predictive* power, in the sense that deciding whether one or the other standard account is right — and therefore whether Johnny is the same person — can affect the way the Johnny is treated, how he is predisposed to think of himself, and so on. The idea concerns a sort of self-fulfilling prophecy: the very *decision to say* whether one of the standard accounts is right can bring about circumstances which supply empirical evidence to retroactively support the view. Call this, for easier reference, the objection of *confirmation by consequences.*

There are many problems with this objection, but I'll mention the salient reason it fails to defeat my argument. The consequences brought about by pronouncing one or another theory of personal identity true fail to fit into the causal structure of explanation and prediction I described in simple terms above. They are, at best, correlative *suggestions* that the explanation or prediction is right.

B.C./2016). The usual method of showing a theory's intuitive appeal is to describe conditions — to propose a thought experiment — under which the theory yields agreeable conclusions, and so appears likely to be right. I pose two challenges to intuition as a theory-choice criterion: first the charge, taking after the intellectual descendants of Rudolf Carnap (1950), that consulting intuition tells us not about the real world, but about the intuiter's conceptual commitments; and second the charge that even if intuition manages to inform us about the real world, the ruling it delivers on personal identity is indeterminate.

We can motivate the first charge thus: say you think, rather reasonably, that *socks aren't people.* Say, further, that we know that proposition to be true. Now, it might be the case that what makes it true is the way that certain metaphysical facts are — viz., the world is such that there's some deep, essential difference between sockhood and personhood. Call this *the metaphysical ground* for your proposition. At the same time, one of the reasons your proposition is true is the fact that our linguistic conventions work in a particular way[34]: the conditions under which we consider it appropriate to label something "sock" are (except in exotic or make-believe cases) utterly different from those under which we consider it appropriate to label something "person".[35] Call this *the semantic ground* for your proposition.

The follower of Carnap (1950) maintains that even though we normally assume that the semantic ground tracks the metaphysical ground — that is, that the distinction between "sock" and "person" reflects some underlying, metaphysical distinction between sockhood and personhood — we can't know whether that underlying distinction

---

[34] Some will contend that this is better regarded as just the satisfaction of assertability conditions, not as truth-making. That's no problem: rephrasing in those terms doesn't damage the basic contention from Carnap (1950) that I lay out. And for an argument that linguistic conventions can make propositions true, see Warren (2015).

[35] To cash this out in more precise terms: let us designate $\{S_1, S_2, S_3, ...\}$ as all the states of affairs (all the possible sets of necessary and sufficient conditions) under which it's appropriate to call something "sock", and $\{P_1, P_2, P_3, ...\}$ as all the states of affairs under which it's appropriate to call something "person". The intuition here is that these sets do not overlap; no S_n is identical to any P_n.

actually exists, *independently of* that linguistic/conceptual scheme that includes the terms "sock" and "person". (Indeed, there are plenty of descriptively superior conceptual schemes — e.g., a fundamental-physical theory of quantum fields — on which no such distinction between sockhood and personhood exists.) In Carnap's terminology, the metaphysical ground can't be known as an "external" matter of fact: asking how things *really* are is asking an unanswerable question. The proposition that socks aren't people is thus evaluable for truth *only* on account of the semantic ground, and not by correspondence with how the world is. So your denial of sockish people might be well informative about the workings of your linguistic/conceptual scheme, interestingly structured as it is to accommodate our workaday perceptual and cognitive tendencies (not to mention our practical interests!), but it cannot tell us how the world really is (Ásta 2013). If this right, then relying on intuition to decide which of the standard accounts is right — as in, say, Locke's mind-swapping experiment (1690/1998; mentioned in §1.2) — can't tell us how things really are. This problem can be summed up in a rather clunky phrase: *nondescriptivism* about intuitive reports.

It may be that you're unconvinced by this neo-Carnapian picture; it may be that you can't shake the idea that there's *some* linguistic and conceptual scheme which we can know to reflect the underlying natural structure of the world. I won't mount an argument here to push back; but if you're so far unmoved, here's the second worry. I am in nominal agreement with Sider (2001) on this issue: PV and BV look evenly matched in terms of their intuitive suitability for ordinary talk of people, making intuition unable to say which of the standard accounts is right.

Here the paradigm demonstration is Williams's "The Self and the Future" (1970), where, on the basis of apparently insignificant factors, intuition equivocates between conclusions — between whether PV or BV is right. Williams presents a scenario in which two people, whose consciousnesses have been swapped into each other's bodies, are about to be tortured. His trick is to tell the same sequence of events in two different ways: first in third-person, as though the torture were meant for others ("Suppose that

there were some process to which two persons, *A* and *B*, could be subjected"), and second in first-person, as though the narrator were one of the people to be tortured ("Someone in whose power I am tells me that I am going to be tortured tomorrow"). Each iteration's telling results in contradictory outcomes: the third-personal version encourages the intuition that something roughly like *PV* is right (a "mentalistic" conclusion, Williams calls it), and the first-personal telling encourages the intuition that something like *BV* ("the bodily continuity identification") is right. Williams concludes:

> The idea . . . that the situation after the experiment is conceptually undecidable in the relevant respect seems not to assist, but rather to increase puzzlement; while the idea (so often appealed to in these matters) that it is conventionally decidable is even worse. Following from all that, I am not in the least clear which option it would be wise to take if one were presented with them before the experiment. I find that rather disturbing.

Should you be inclined to deny my neo-Carnapian charge of nondescriptivism but find plausible this problem of intuitive undecidability, you might yet hold out for the success of some future thought experiment at rendering a verdict decisively in favor of one theory. But as things stand, we have little reason to believe that intuition does or will give us any such verdict.

Here's our last type of consideration: the theoretical virtues. The theoretical virtues include parsimony — keeping the number of entities a theory postulates to a minimum — and elegance — accounting for as wide a range of phenomena as possible with as few principles as possible — among others (Willard 2013). So by the lights of the theoretical virtues, standard formulations of *BV*, which define personal identity over time as *just* a matter of having a continuously living body, might be judged preferable over a version of *PV* that demands a great number of mental features — say,

"memories, intentions, beliefs/goals/desires, and similarity of character" (Parfit 1984: 207) — to sustain personal identity over time.

Compact, simple statements of scientific theories often garner admiration for their 'beauty'.[36] But it remains to be shown that the compactness or simplicity of *metaphysical* theories would make them likely to be *true*.[37] Willard (2013) makes the case that "arguments for simplicity in the natural sciences cannot be easily transformed into arguments for simplicity in metaphysics, and that in order to preserve simplicity as a metaphysical criterion, we need to provide independent motivation to accept simplicity on metaphysical grounds." So what the theoretical virtues may do for scientific theories they plausibly cannot do for metaphysical theories: track the likelihood of their truth. Thomasson (2014a) suggests a reason why contemporary philosophers might have believed (and in certain cases, still do believe (Bradley 2017)) otherwise: having been impressed by the tremendous successes of theory construction in the natural sciences, they came to think that emulating the formal features of good scientific theories would shore up their own theoretical work.[38] Whatever the case may have been, it's plausible that the theoretical virtues improve the *aesthetic appeal* of metaphysical theories such as the standard accounts of personal identity — but not their likelihood of being true.

---

[36] See, for instance, Farmelo (2002): "For [the physicist Steven] Weinberg . . . Einstein's theory of gravity is an exemplar of a beautiful scientific theory . . . based on a few simple principles."

[37] W. V. O. Quine (1966) suggests a few reasons. Among the stronger ones are the greater ease of confirming a (more parsimonious) theory that posits fewer entities whose existence needs to be proved; and the greater ease of confirming a (more elegant) theory that can account for a wider range of natural events, since the greater number of accounted-for events increases the theories' chances of being confirmed. But quite contrary to the inclinations of philosophers engaged in the standard disputes over personal identity, these two reasons evince a peculiar insouciance about what actually makes theories right: they tell us that more parsimonious and elegant theories are likelier to be true only because they were formulated cleverly enough to make confirmation an easier task. If philosophers who think of themselves as serious truth-seekers want to know why a theory's simplicity could be reliably truth-tracking, they won't be convinced by Quine's reasons. Also, see Huemer (2009).

[38] None of this is to say, of course, that *something like* the modern theoretical virtues hadn't been offered before; Ptolemy is said to have advocated an early form of theoretical simplicity.

If this is right, then we're at the end of our rope. The three types of theory-choice criteria — empirical adequacy, intuition, and the theoretical virtues — have all failed to tell us which of the standard accounts is likelier to be right. This failure means, strange through it may seem, that we can't say *what* the metaphysical facts about people are. This already suffices to motivate the "epistemic diagnosis" advanced in §1.3 — that we have little to no justification for believing that the answers to the questions of personal identity are one way or another — but we can motivate an even stronger resignation. For the three structural problems afflicting each theory-choice criterion — empirical equivalence, nondescriptivism about intuitive reports, and the plausible incapacity of theoretical virtues to track the truth — apply not just to the debate between the advocate of *PV* and the advocate of *BV*, but also, it turns out, to the debate between the *realist* and the *nihilist* about people. More trouble awaits the philosopher of personal identity.

3.3. *Between realism and nihilism: knowing whether there are facts*

Whereas the issues that divide defenders of *PV* and *BV* disagree are personhood and persistence, the question on which realists about people and nihilists about people (henceforth shortened to just *realist* and *nihilist*) disagree is

EXISTENCE: Are there people?[39]

The realist says *yes, obviously!* The nihilist says *no,* people don't *really* exist; all there are are collections of simples (whatever the most fundamental units of matter turn out to be on our best physical theory: particles, fields, superstrings, etc.) arranged like a person, or "person-wise" (van Inwagen 1990). Recall that what we established in §3.2

---

[39] See fn 2 for a semi-formalization.

was undecidability between PV and BV, with the consequence that we can't say what the metaphysical facts about people are. What I aim to establish here is undecidability between realism and nihilism, with the more harrowing consequence that we can't know *whether* there are metaphysical facts about people at all. To do so, we simply need to reapply our three theory-choice criteria — and observe their failure.

First: the question of empirical adequacy. Take a toy case: suppose a realist and a nihilist are arguing over whether their mutual friend Uri is a person. The realist seems astonished that the question even comes up, while the nihilist avers that in fact, Uri is nothing more than a person-wise composite of quantum fields. To ask which side empirical adequacy favors, we ask which side better explains or predicts the observational data. Just as in the dispute (in §3.2) over Johnny the coma patient's persistence, it appears here that neither of the two theories gives a better explanation or prediction of Uri's ontological status (his being or not being a person, as the case may be). For the realist maintains a simple thesis — that Uri is a person — and the nihilist maintains an simple thesis — that Uri is not a person, but just a quantum-field aggregate — and yet it seems difficult, bizarre even, to ask whether evidence gained from observing Uri better fits the realist thesis or the nihilist thesis, since it generally appears to fit both *equally well.* Indeed, I sense that the tempting thing to say here is that we really ought to trade off the putative metaphysical dispute for a semantic one: to say that Uri is a person and to say that Uri is a quantum-field aggregate is just to provide two different *descriptions* of Uri corresponding to two different scales of analysis, and not to substantively disagree about what Uri in fact is. Even if this temptation is misguided — and I think it in fact is, as we'll see in §4 — the equal fit of both theories with the empirical evidence means that just like PV and BV, realism and nihilism suffer from empirical equivalence.

How about intuition? Here we need not provide a particular thought experiment, because if the charge of nondescriptivism that I developed in the last section — that intuition informs us not about the real world, but merely about the intuiter's

conceptual commitments — is right, then intuition's problem is *global*: there are no cases in which accordance with intuition is a reliable theory-choice criterion. So despite the realist's knee-jerk affirmation of common sense, it remains to be shown that what he asserts — that *of course* there are people — is anything more than mere chauvinistic allegiance to a linguistic/conceptual scheme that has served him well. (One can imagine, for fun, a nihilist who comes along and informs the realist that her *own* lifelong training in elementary particle physics has cultivated in her a rather different set of intuitions about how the world really is: that every apparently solid thing in the world, including every person, is just a smearing of points in Minkowskian spacetime. Without appeals to further virtues like empirical adequacy and conceptual coherence, I cannot see how the knee-jerk realist could reasonably respond to her.)

Finally: the theoretical virtues. Presumably, the nihilist has the *prima facie* upper hand here, since she requires fewer explanatory posits than does the realist (at least the realist who accepts the standard theories in contemporary physics). Whereas the scientifically well-informed realist about people is forced to maintain that there are at least two kinds of entity in the world — people *and* more fundamental building blocks of matter (e.g., quantum fields) — the nihilist holds fast to the austere fundamental-physical picture, on which only simples exist. But the nihilist's parsimony plausibly doesn't buy him any advantages when it comes to his theory's likelihood of being true. As is the case with intuition, so it is with theoretical virtues: the trouble is global. If it's true, as I suggested in §3.2, that theoretical virtues fail to track the truth — and indeed, we have no demonstration that they do — then theoretical virtues cannot tell us whether realism or nihilism is likelier to be right.

What's left for us is to draw out the conclusion of undecidability between the realist and the nihilist theses. First, to spell out their respective views:

REALISM: There exists at least one x such that x is a person.

42

NIHILISM: There does not exist an x such that x is a person.

The epistemic consequence is straightforward: if we have no way to tell which of these is in fact right, we have no way to tell whether there ('really', metaphysically) exist people. And if we have no way to tell whether there exist people, the threat of catastrophic presupposition failure (for description, see under the "nihilist diagnosis" in §1.3) looms large over any and all claims about personal identity. So we are left unable to say not just *what* the metaphysical facts about people are, but *whether* there are any at all.

What I want to suggest, finally, is that the twofold metametaphysical failure just witnessed — first in the dispute between the defender of PV and the defender of BV, and second in the dispute between the realist and the nihilist about people — is not an accident. Matters of personal identity have a peculiar character that makes them prone to this sort of failure. It is plausibly on account of, or at least partly on account of, the *derivative* and *causally inert* nature of personal identity that candidate theories about any aspect of personal identity are susceptible to epistemic problems like undecidability. What do I mean? First, aspects of personal identity — personhood, persistence, and existence — are said to obtain in virtue of facts that are more fundamental; they *derive from,* or *are grounded in* them. Second, these derivative properties — being a person, being the same person over time — exhibit *no causal powers* over and above those already exhibited in its grounding facts, and so generate no observable data to justify the postulation of their existence. This is plausibly why our asking about Johnny the coma patient and about Uri the mutual friend seemed fruitless: the causally inert character of Johnny's retained (or lost) persistence and the causally inert character of Uri's personhood (or nonpersonhood) were such that no new, independent data would be available to favor either side's thesis in the two debates. Thus do attributions of personhood and persistence end up as mere metaphysical tags,

markers that do nothing except to point to those facts from which they're derived.[40] What results is a general metametaphysical trouble that, with enough surveying, we start to notice in many other domains of metaphysical inquiry: it appears nearly impossible to come up with a reasonable epistemology for any theory that posits derivative and causally inert entities or properties.[41]

4. The semantic diagnosis: are the disputes "merely verbal"?

That completes my case for what I've called the "epistemic diagnosis". Before turning to a positive account of personal identity, I want to entertain an alternative proposal for making sense of "the fuss", also suggested in §1.3, that looks superficially attractive: the *semantic diagnosis*. What if the source of all the metametaphysical trouble with personal identity is not epistemic opacity, but linguistic confusion?

On a first gloss, the idea of the semantic diagnostician is this: the real trouble is that while the disputes appear to be genuinely substantive — i.e., involving distinct, rival pictures of how things are — they are in fact "merely verbal" — merely concerned with what terminology we prefer for talking about things. This sort of mistake could happen in at least two ways: (1) when two disputants utter the same sentence (or semantically similar sentences) but take the sentence to refer to different and seemingly incompatible states of affairs; or (2) when disputants take different and seemingly contradictory sentences to refer to the same state of affairs. On a charitable understanding of ordinary linguistic behavior, (1) typically comes out as the rather uninteresting phenomenon of *ambiguity*, but it covers any situation in which (1')

---

[40] The neo-Austinian objection — what I called in fn 32 the confirmation by consequences — can be tempting here. See fn 32 for a response.

[41] I set forth this soft kind of verificationism rather unabashedly, thanks to recent strides made by neo-Carnapians; see especially Warren (2017) for a snappy recent argument against what he terms "non-causal realism".

disputants utter the same sentence but take it to have different satisfaction conditions or truth-conditions. Thus, in a nice example from Karen Bennett (2009), the sorority girl who indiscriminately considers any old drink served in a V-shaped glass to be a martini takes the sentence "That's a martini" to have the sufficient truth-condition of there being any old drink in a V-shaped glass — that is, to be true as long as there's any old drink in a V-shaped glass. The martini purist, however, considers "That's a martini" only to be satisfied by a state of affairs in which there's a "cocktail ... made of gin or vodka, dry vermouth, and perhaps an olive or two." So if a V-shaped glass of apple liqueur is in front of them, the sorority girl and the purist will *appear* to have an ontological disagreement: the sorority girl will claim that there exists a martini, while the purist will say that there does not. This kind of definitional disagreement, whose vacuousness goes unnoticed both by the sorority girl and by the purist, is, the semantic diagnostician wants to say, what is really going on in many serious-sounding metaphysical discussions.

Let's sharpen up the notion of a dispute's being "merely verbal". Eli Hirsch (2005: 83) says we can tell that a dispute is merely verbal if "there are two undisputed sentences U_1 and U_2, one true and one false, such that one side holds that [a disputed sentence] D is (*a priori* necessarily) equivalent to U_1 and the other side holds that D is equivalent to U_2". Call this Hirsch's schema. What's more, if the dispute is in fact merely verbal, we should be able to resolve it happily by implementing a translation scheme: "each side ought to acknowledge that there is a plausibly charitable interpretation of the language associated with the other side's position which will make that position come out true" (Hirsch 2005: 82). So: might this be the right picture — a picture that's explanatorily superior to that of the epistemic diagnosis — of what is going on when metaphysicians are arguing, seemingly in vain, over personal identity?

I'm fairly confident that the answer is *no*. For it doesn't appear that despite its careful formulation, Hirsch's schema works consistently across disputes — in particular, on those all-important disputes over personhood and persistence. Even more

45

worryingly, it doesn't appear to work on precisely the sort of ontological disagreements it was engineered to resolve — in this case, the existence dispute between the realist and the nihilist. It happens to deal with Bennett's martini case just fine (Bennett 2009: 14), but as Bennett admits, the martini case is special — unusually well-adjusted for Hirsch's schema. Let's try out Hirsch's schema on the disputes we're interested in.

Between the realist and the nihilist, the disputed sentence D is "There are people." Most realists who accept some standard theory or other in contemporary physics will concur with the nihilist that *there are simples arranged person-wise*, so let us try designating that as the undisputedly true sentence U_1 maintained by the nihilist. So far, so good: it's plausible that D and U_1 are *a priori* and necessarily equivalent, and the scientifically well-informed realist takes both D and U_1 to be true. But what, now, is the undisputedly false sentence U_2 that the realist is supposed to maintain? It can't be "There are people," because that's already the disputed sentence D and it's not undisputedly false. Nor can it be "There are no people," because the realist would not maintain such a claim (and it's also not undisputedly false). Suppose we attempt a different way to fill out the placeholders. Say that D = "There are no people," that U_1 = "There are simples arranged person-wise," and stipulate (for charity's sake, if anything) that D and U_1 are equivalent. Once again, what is U_2 supposed to be? It still cannot be "There are people," since that is disputed by the nihilist and it's not undisputedly false. Nor can it be "There are no people." And considering that the stock of available sentences that each side can maintain is extremely limited, we seem to have no plausible permutations for sentence-designation left. Thus unable to complete Hirsch's schema, I simply can't see how we can allege any precise equivalence — attractive though it may be from a distance — between the realist and the nihilist theses. This should already give us some pause: the dispute between the realist and the nihilist is not "merely verbal" (in the Hirschian sense).

Let's turn now to the dispute between the defenders of the standard accounts. Filling out Hirsch's schema is even more difficult here. For the disputed sentence D,

we can opt for a statement of personhood like "Personhood obtains in virtue of having certain psychological features", taken straight out of the PV playbook. But we soon find that D is not a plausible candidate for *a priori* and necessary equivalence with any of the undisputed sentences we could provide for the defender of PV or the defender of BV. For instance, the advocate of PV could assert, "There are psychological features", without contest from her opponent. (Notice that we'll encounter trouble if we try to buff up the statement; a sentence like "There are psychological features *that matter for personhood*" will be contested by the defender of BV.) Or she could appeal to a foundational premise on which she agrees with the advocate of BV: "Personhood obtains in virtue of having certain essential features." Now, if we designate either "There are psychological features" or "Personhood obtains in virtue of having certain essential features" as U_1, we immediately encounter a problem: U_1 clearly won't be *a priori* and necessarily equivalent to D. In the case of "There are psychological features", it isn't even in the ballpark of inference. Other arrangements will not satisfy Hirsch's schema, either. The advocate of BV could say, "We have bodies," or, just for kicks, "Persistence obtains in virtue of certain essential features." But no conceivable disputed sentence D appears to be *a priori* and necessarily equivalent to either of these statements.

What's the upshot of this tedious procedure? At least on the most careful conception I know of a dispute's being merely verbal — the Hirschian one — the disputes over personhood, persistence, and existence are not merely verbal. Indeed, if we accept the epistemic diagnosis I made in §3, this denial of the semantic diagnosis should be thoroughly plausible. The disputes themselves may well be perfectly substantive, and well-maneuvered around linguistic pitfalls; it's just that we can't, for *epistemic* reasons, adjudicate which side in any of the disputes is right. This can wind up rather confusing. But we're not about to back down just yet. One more step: I want to propose that even as the *disputes* over the metaphysics are not merely verbal, we would do best to regard the first-order metaphysics of people — matters of personal

identity — as something akin to a grand fiction, tales and conventions spun by our ordinary linguistic behavior.

5. What should we say?

If you've come with me this far, it can seem rather unusual that we go about our ordinary lives just fine without knowing the truth about personal identity. But no mystery is at work here. Our ordinary claims about personal identity, you'll recall from §2, are presuppositions with purposes — indispensable fictions which can be sustained, and which we have obvious reasons to sustain, regardless of whether there are facts of matter about people.[42] These fictions — these shifty, stipulative concepts whose meanings change with our varying social purposes — are what's left.

My suggestion is simple. Rather than trying to sleuth out the deep metaphysical nature of people — a grand philosophical project that has, by its own lights, gone just about nowhere — we should accept the ordinary concept of "person" as *good enough*. This is *pragmatism* about personal identity, tempered and encouraged by the awareness that we don't have the resources to answer the questions of personal identity.[43] So what would an account of personal identity based on our ordinary concept of "person"

---

[42] To be clear: my use of the word "fiction" isn't meant to suggest that I favor a *fictionalist* view of personal identity. I won't have the chance here to detail how the pragmatic account I favor differs.

[43] This is meant as a nod to the American and British traditions of pragmatism that emerged with C. S. Peirce (1878), William James, and F. P. Ramsey (Tiercelin 2004), who, despite their differences, held in common a view of truths as *beliefs that work*. Here's James in *The Meaning of Truth* (1909/1997):

> *The true,* to put it very briefly*, is only the expedient in the way of our thinking, just as the right is only the expedient in the way of our behaving.* Expedient in almost any fashion, and expedient in the long run and on the whole, of course . . . . Experience, as we know, has ways of *boiling over*, and making us correct our present formulas. [italics in the original]

I should note that that whereas James was a staunch *global* pragmatist, I'm a reserved *local* pragmatist: I don't, unlike James, think that we can justify the reduction of whether a theory is true to whether it's useful in *all* domains of inquiry.

look like? It would have to be flexible and general enough to accomodate the inconsistent and changing ways we talk about people across social, legal, political, and other contexts. Here's a shot at formulations for personhood and persistence:

LOCAL PRAGMATISM ABOUT PERSONHOOD: Relative to any context C and any time ϕ, x is a person if and only if the predicate "is a person" is applicable to x at $\langle C, \phi \rangle$.

LOCAL PRAGMATISM ABOUT PERSISTENCE: Relative to any context C and any time ϕ, x at t_1 and y at t_2 are the same person if and only if the predicate "is the same person" is applicable to $\langle x, y \rangle$ at $\langle C, \phi \rangle$.

On a view like this, we drop the search for essential features, admitting instead that all that makes people *people* are meaningful contexts: those jungles of social norms and linguistic rules that shape and stabilize the meanings of words and concepts. That is: matters of personal identity are, fundamentally, nothing more than matters of language use.[44] But if the pragmatic account is right, personal identity has been really been a semantic phenomenon *all along*, a curious effect of our ordinary language, intuitions, and practices — of the ways we go about our daily lives. And rather appealingly, viewing the metaphysics of people in this sober way helps us to keep in mind what's at stake here: the fact of our meaning-determining practices makes it ultimately *up to us* what does and doesn't count as a person, what counts as the same person over time, and so on. This matters. Whether or not, for instance, we attribute personhood to corporations (Rovane 1998; *Hobby Lobby v. Sebelius* 2013), to fetuses (*Roe v. Wade* 1973), to AI (Solum 1992; Schwitzgebel & Garza 2015), to the severely cognitively

---

[44] Quine (1980) named this kind of higher-order move, where one reframes metaphysical questions as linguistic questions, *semantic ascent*. See Wittgenstein (1953/1968) and Carnap (1950) for the historical inauguration of the method in deflationary metaphysics, and Price (forthcoming) for an excellent recent defense.

disabled (McMahan 1996; Kittay 2008), to non-human animals (Feltman 2014), or to any other not-so-easy cases has great social repercussions.

One question remains. If after centuries of clever argumentation, the standard disputes over personal identity have failed to yield a conclusion, what are we to make of philosophers who insist on carrying on the debate? We might resolve to say that their efforts are moot; but a more charitable and promising way to view what they're doing is in terms of "metalinguistic negotiation": normative disagreement over how the concept and the word "person" *ought* to be used (Plunkett & Sundell 2013; Plunkett 2015; Thomasson 2016). So reconceived, disputes over personal identity are disputes in "conceptual ethics": forms of collective deliberation about how we want our concepts to work for us, what kinds of distinctions we want to draw in the world, what kinds of rights, responsibilities, and other statuses we want to assign to others, and so on.[45] And if that's the case, plenty of work awaits us.

---

[45] See Burgess & Plunkett (2013a,b) for a sustained defense of the field.

References

Baker, Lynne Rudder. (2002). The Ontological Status of Persons. *Philosophy and Phenomenological Research*, 65, 370-388.

Bennett, Karen. (2009). Composition, Colocation, and Metaontology. In D. Chalmers, D. Manley, & R. Wasserman (eds.), *Metametaphysics.* Oxford: Oxford University Press.

Bennett, Karen. (2011). By Our Bootstraps. *Philosophical Perspectives,* 25 (1), 27-41.

Berker, Selim. (forthcoming). The Unity of Grounding. *Mind.*

Bostrom, Nick. (2003). Are we living in a computer simulation? *Philosophical Quarterly,* 53 (211), 243–255.

Bourget, David & Chalmers, David. (2014). What do philosophers believe? *Philosophical Studies,* 170 (3), 465-500.

Bradley, Darren. (2017). Philosophers should prefer simpler theories. *Philosophical Studies*, 1-19.

Burgess, Alexis & Plunkett, David. (2013a). Conceptual Ethics I. *Philosophy Compass,* 8 (12), 1091-1101.

Burgess, Alexis & Plunkett, David. (2013b). Conceptual Ethics II. *Philosophy Compass,* 8 (12), 1102-1110.

Burke, Michael. (2003). Is my head a person? In K. Petrus (ed.), *On Human Persons.* Frankfurt: Ontos.

Campbell, Tim & McMahan, Jeff. (2010). Animalism and the varieties of conjoined twinning. *Theoretical Medicine and Bioethics,* 31 (4), 285-301.

Carey, Susan. (2011). Précis of 'The Origin of Concepts'. *Behavioral and Brain Sciences*, 34 (3), 113-24.

Carnap, Rudolf. (1950). Empiricism, Semantics, and Ontology. *Revue Internationale de Philosophie*, 4, 20-40.

Chalmers, David. (2011). Verbal Disputes. *Philosophical Review,* 120 (4), 515-566.

Farmelo, Graham. (2002, February 20). Beautiful equations to die for. *The Telegraph.* Retrieved from *http://www.telegraph.co.uk/.*

Feltman, Rachel. (2014, December 22). Orangutan granted rights of personhood in Argentina. *The Washington Post.* Retrieved from *https://www.washingtonpost.com/.*

Gendler, Tamar Szabó. (2002). Personal Identity and Thought-Experiments. *Philosophical Quarterly,* 52 (206), 34-54.

Hershenov, Daniel. (2006). Shoemaker's Problem of Too Many Thinkers. *Proceedings of the American Catholic Philosophical Association,* 80, 225-236.

Hershenov, Daniel. (2013). Who Doesn't Have a Problem of Too Many Thinkers? *American Philosophical Quarterly,* 50 (2), 203-208.

Hirsch, Eli. (2002). Quantifier Variance and Realism. *Philosophical Issues,* 12, 51-73.

Hirsch, Eli. (2005). Physical-Object Ontology, Verbal Disputes, and Common Sense. *Philosophy and Phenomenological Research*, 70, 67-97.

Hobby Lobby Stores, Inc. v. Sebelius, 723 F. 3d 1114 (10th Cir. 2013)

Huemer, Michael. (2009) When is Parsimony a Virtue? *The Philosophical Quarterly,* 59 (235), 216-236.

Hume, David. (2000). *A Treatise of Human Nature.* Oxford: Oxford University Press. (Original work published 1739)

James, William. (1997). *The Meaning of Truth.* Prometheus Books. (Original work published 1909)

Kittay, Eve F. (2008). At the margins of moral personhood. *Ethics,* 116, 100-131.

Kleene, Stephen. (2009). *Introduction to Metamathematics.* New York: Ishi Press. (Original work published 1952)

Kriegel, Uriah. (2013). The Epistemological Challenge of Revisionary Metaphysics. *Philosophers' Imprint,* 13 (12), 1-30.

Locke, John. (1998). *An Essay Concerning Human Understanding.* London: Penguin Classics. (Original work published 1690)

Lowe, E. J. (1995). Coinciding objects: in defence of the 'standard account'. *Analysis,* 55 (3), 171-178.

McMahan, Jeff. (1996). Cognitive Disability, Misfortune, and Justice. *Philosophy & Public Affairs,* 25, 3-35.

Merricks, Trenton. (1998). There Are No Criteria of Identity Over Time. *Noûs*, 32, 106-124.

Olson, Eric. (1997a). *The Human Animal: Personal Identity Without Psychology.* Oxford: Oxford University Press.

Olson, Eric. (1997b). Was I Ever a Fetus? *Philosophy and Phenomenological Research,* 57 (1), 95-110.

Olson, Eric. (2001), Material Coincidence and the Indiscernibility Problem. *The Philosophical Quarterly,* 51, 337-355.

Olson, Eric. (2003). An argument for animalism. In J. Barresi and R. Martin (eds.), *Personal Identity.* Oxford: Blackwell.

Paul, L. A. (2012). Metaphysics as modeling: the handmaiden's tale. *Philosophical Studies,* 160 (1), 1-29.

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

Peirce, C. S. (1878). How To Make Our Ideas Clear. *Popular Science Monthly,* 12, 286-302.

Plato. (2016). *The Republic of Plato.* (Allan Bloom, trans.). New York: Basic Books. (Original work published ca. 380 B.C.)

Plunkett, David. (2015). Which Concepts Should We Use?: Metalinguistic Negotiations and The Methodology of Philosophy. *Inquiry,* 58 (7-8), 828-874.

Plunkett, David & Sundell, Tim. (2013). Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint,* 13 (23), 1-37.

Price, Huw. (forthcoming). In I. Ravenscroft (ed.), *Minds, Worlds & Conditionals: Essays in Honour of Frank Jackson.* Oxford: Oxford University Press.

Puccetti, Roland. (1973). Brain Bisection and Personal Identity. *The British Journal for the Philosophy of Science,* 24 (4), 339-355.

Quine, W. V. O. (1966). On Simple Theories of a Complex World. In Quine, *The Ways of Paradox.* Random House.

Quine, W. V. O. (1980). *From a Logical Point of View.* Cambridge: Harvard University Press.

Roe v. Wade, 410 U.S. 113 (1973)

Rosen, Gideon. (2010). Metaphysical Dependence: Grounding and Reduction. In B. Hale & A. Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology.* Oxford: Oxford University Press.

Rovane, Carol. (1998). *The Bounds of Agency.* Princeton: Princeton University Press.

Schaffer, Jonathan. (2009). On What Grounds What. In D. Chalmers, et al. (eds.), *Metametaphysics.* Oxford: Oxford University Press.

Schwitzgebel, Eric & Garza, Mara. (2015). A Defense of the Rights of Artificial Intelligences. *Midwest Studies in Philosophy*, 39, 98-119.

Shoemaker, Sydney. (1991) Personal Identity: A Materialist's Account. In S. Shoemaker & R. Swinburne, *Personal Identity.* Oxford: Blackwell.

Sider, Theodore. (1996). All the World's a Stage. *Australasian Journal of Philosophy,* 74, 433–453.

Sider, Theodore. (2001). Criteria of personal identity and the limits of conceptual analysis. *Philosophical Perspectives,* 15, 189-209.

Solum, Lawrence. (1992). Legal personhood for artificial intelligences. *North Carolina Law Review*, 70, 1231.

Sveinsdóttir, Ásta. (2013). Knowledge of essence: The conferralist story. Philosophical Studies, 166, 21-32.

Thomasson, Amie. (2013). The easy approach to ontology: a defense. In M. Haug (ed.), *Philosophical Methodology: The Armchair or the Laboratory?* Routledge Press.

Thomasson, Amie. (2014a). *Ontology Made Easy.* Oxford: Oxford University Press.

Thomasson, Amie. (2014b). Quizzical ontology and easy ontology. *Journal of Philosophy,* CXI (9/10), 502-528.

Thomasson, Amie. (2016). Metaphysical Disputes and Metalinguistic Negotiation. *Analytic Philosophy*, 57 (4), 1-28.

Tiercelin, Claudine. (2004). Ramsey's Pragmatism. *Dialectica, 58* (4), 529-547.

Unger, Peter. (1979a) I do not exist. In G. F. MacDonald (ed.), *Perception and Identity.* London: Palgrave.

Unger, Peter. (1979b). Why There Are No People. *Midwest Studies in Philosophy*, 4 (1), 177-222.

van Inwagen, Peter. (1990) *Material Beings.* Ithaca: Cornell University Press.

von Fintel, Kai. (2004) Would you believe it? The king of France is back! Presuppositions and truth-value intuitions. In Reimer, M. & Bezuidenhout, A. (eds.), *Descriptions and Beyond.* Oxford: Oxford University Press.

Warren, Jared. (2017). Epistemology versus non-causal realism. *Synthese,* 194 (5).

Wittgenstein, Ludwig. (1968). *Philosophical Investigations.* Oxford: Blackwell. (Original work published 1953)

Willard, M. B. (2013). Game called on account of fog: metametaphysics and epistemic dismissivism. *Philosophical Studies* 164 (1), 1-14.

Williams, Bernard. (1970). The Self and the Future. *Philosophical Review*, 79 (2), 161-180.

Yablo, Stephen. (2006). Non-Catastrophic Presupposition Failure. In A. Byrne and J. Thomson (eds.), *Content & Modality.* Oxford: Oxford University Press.

Yablo, Stephen. (2009). Must Existence-Questions Have Answers? In D. Chalmers, et al (eds.), *Metametaphysics.* Oxford: Oxford University Press.

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