***Animal Interrupted,***

***or Why Accepting Pascal’s Wager Might Be the Last Thing You Ever Do***

**1. What Our Paper Is About**

It might seem a bit odd to give a paper about conventionalist views of personal identity at a conference on animalism, given that (as we’ll discuss in more detail below) the two views in many ways appear to address different sets of issues and solve different sorts of problems. Animalism’s main competitor is usually taken to be Lockeanism (which denies that we are identical to organisms and claims that psychological continuity of some sort is what’s necessary and sufficient for our persistence).[[1]](#footnote-1) Conventionalism, by contrast, doesn’t agree *or* disagree with animalism per se. Rather, it offers itself as a radical alternative to traditional accounts of personal identity that can accommodate both animalist and Lockean intuitions; on conventionalism, persons are constituted in part by practices and attitudes of certain sorts of care, and both physical organisms and psychological entities can be the subject of such practices and attitudes.[[2]](#footnote-2)

What’s more, conventionalism -- at least the robust version we consider in this paper -- is not a particularly popular view in personal identity circles, and so it might seem a bit of an obscure target. Yet, as social ontology, social epistemology, and anti-realist views all continue to gain philosophical traction, conventionalism about personal identity is increasingly (if often tacitly) accepted. In this paper, we concentrate on the most well-developed and defended version of conventionalism currently on offer (namely, that proposed by David Braddon-Mitchell, Caroline West, and Kristie Miller) and discuss how the conventionalist appears forced either 1) to accept arbitrariness concerning from which perspective to judge one’s survival or 2) to maintain egalitarianism at the cost of making ‘transfiguring’ decisions such as Pascal’s Wager rationally intractable. We consider three ways the egalitarian conventionalist could make these choices tractable and show that each one comes at significant cost to the view. We end the paper by considering whether accepting arbitrariness would be a better move for the conventionalist and conclude that even here, she runs the risk of transfiguring choices being rationally intractable.

**2. Identity (and Survival) by Convention?**

One of the main objections to animalism is that it can’t accommodate ‘gappy’ or ‘jumpy’ existence -- or, really, the idea of post-mortem survival in general.[[3]](#footnote-3) If continuity of organic processes of a certain sort is necessary for our survival, then stepping into a teleporter means our death, and even God can’t resurrect an animal identical to the one(s) currently writing this paper.[[4]](#footnote-4) Conventionalism is championed by its advocates, by contrast, as an alternative view that can handle both these and other ‘puzzle’ cases, such as those involving fission, fusion, and brain transplants. In this section, we lay out the bare bones of this view, showing how the most developed form of conventionalism answers the question of *(a)* what persons are and *(b)* what the necessary and sufficient conditions for their persistence are.

*(a) what persons are*

Conventionalism about personal identity has a long and storied history, going back at least to Hume, who in his *Treatise on Human Nature* compared persons to commonwealths or nations.[[5]](#footnote-5) In more recent years, it has been championed by the likes of Parfit, Unger, West, Braddon-Mitchell, and Miller.[[6]](#footnote-6) There are a number of variations on the basic view, but all versions agree that personal identity is tied up with personal and/or social conventions in one way or another.[[7]](#footnote-7) The version of conventionalism on which we will focus in this paper takes the relationship between personal identity and conventions to be one of constitution. That is, for the strong conventionalist, persons are partly constituted by conventions of a certain kind.

To get a basic sense for the view, consider Hume’s analogy of a nation. A nation is constituted by 1) a land-mass along with 2) people who live (primarily) on that land-mass and 3) conventions regarding the rules of association for persons within the boundaries of the relevant land-mass. Strip the nation of any of these three elements and, arguably, you remove the nation from existence. Strong conventionalism about persons treats persons like nations. Persons, on this view, are (at a minimum) constituted by some physical stuff - a body - plus some psychological features and a range of conventions.[[8]](#footnote-8)

At first glance, this view might seem...well, crazy. How could it possibly be the case that whether or not this room contains persons is a matter of convention? Here it is important to note that the term ‘convention’ in this account doesn’t have the sense of ‘*mere* convention’ in the Lewisian sense (i.e., matters that are malleable, that need to be coordinated, but where there is no requirement around which coordination takes place).[[9]](#footnote-9) The conventions that the strong conventionalist has in mind are really practices and attitudes of a certain kind: namely, practices and attitudes of care. In particular, these person-directed practices include moral and prudential concerns, such as self-regard (including planning) for future continuers and self-attribution of praise or blame; they also include what appear to be social conventions such as when someone is held accountable for murder in a court of law.

Even granting that the conventions involved are practices and attitudes of care, however, there is still the worry that “...when it comes to persons, as opposed to when it comes to, say, nations and clubs, there is a further fact about how we *should* structure our concerns, meaning that how we should structure our concerns is not a mere conventional matter.” (Eklund 2004, added emphasis) That is, it seems that the object of our person-directed practices can’t just be arbitrarily selected -- and, furthermore, that those practices shouldn’t be able to shift in the way that conventional concerns do. We shouldn’t, for example, be able to structure our concerns around rocks and be able to claim truly that those rocks are persons. Nor should we be able to structure our person-directed practices around rocks now and then shift to structuring those practices around lakes in six months.

The conventionalist will respond to this concern by pointing out that this, again, confuses the relevant conventions with *mere* conventions. While it is possible to change *mere* conventions on a whim (and, e.g., place ‘north’ on the bottom of a compass or map), person-directed practices are not of this kind. Strong conventionalists go so far as to say that some of the relevant conventions (such as self-regard for future continuers) are hard-wired into us via evolution, and so may not be malleable at all.[[10]](#footnote-10)

To get a better sense for the general view, consider how conventionalism handles the classic transporter case. Person A enters a transporter at time *t* at location *L*, at which point A’s molecules are destroyed and a distinct set of molecules is created at the output terminus of the transporter at location L\*. This set of of molecules is psychologically continuous with person A, but is not physically continuous with A.

Has A survived the transporter event? Proponents of the psychological continuity view say **‘yes’**: the thing that leaves the transporter is psychologically continuous with the thing that entered, and that’s sufficient for survival. Proponents of animalism and of the physical continuity view say **‘no’**: the entity that leaves the transporter is not physically continuous with A and thus not the same animal. The conventionalist claims that **both** groups are right and that this is a no-fault disagreement. Why? Because the two groups contain distinct *kinds* of persons: the psychological community contains persons that are constituted partly by person-directed practices geared toward psychological continuity, and the physicalist community contains persons that are constituted partly by person-directed practices geared toward physical continuity. According to strong conventionalism, all of these persons exist: some survive transportation, and some do not. Many of the conventions that constitute persons are ‘settled’; that is, they are held in common across various communities. But there are some conventions that are not settled in this way. The conventionalist believes that this is the moral of the puzzle cases of personal identity: in some cases, our conventions diverge and, in so doing, reveal to us the existence of distinct (though perhaps overlapping) communities of persons.

Strong conventionalists do permit the possibility of changes in the conventions that are constitutive of personhood, however, and so -- even if it is not a completely arbitrary matter -- personhood can seem to be a rather transient, unstable thing. At the very least, it seems odd to say that whether or not one survives stepping into a transporter or physical death depends on what community one is a part of, or what one’s personal beliefs are. But to point out that strong conventionalism doesn't seem intuitive on these grounds isn't an objection *per se*; the conventionalist will respond: “That’s not an objection; that’s the view!”[[11]](#footnote-11) The whole point of conventionalism is to offer a radical new way of rethinking the debate surrounding personal identity. As Kristie Miller puts it, “On this view persons are multiply realisable. ‘Person’ is a term that picks out whatever it is that plays some functional role—roughly, the role of being the object around which our person-directed practices are structured. Since different persisting objects can realise that role, which objects in a world are the persons will depend, in part, on how the person-directed practices are structured in that world.”[[12]](#footnote-12) If this is true, we should not be surprised (if still amazed) to discover that two beings who appear from the outside to be identical (e.g., monozygotic twins) can have different persistence conditions.

In the rest of this paper, we challenge the conventionalist not on grounds of absurdity but on a different front, arguing that this view forces her to choose between two unpalatable options: 1) arbitrariness concerning from which perspective to judge survival or 2) the intractability of decisions like Pascal’s Wager. To make this case, however, we need to say more about the conventionalist’s account of survival and persistence.

*(b) persistence conditions*

The strong conventionalist offers a simple initial account of survival: "To survive...is to preserve whatever property a person's (or perhaps community's) person-directed practices are organized around."[[13]](#footnote-13) In other words, a person persists so long as she possesses the central property/ies (psychological or physical continuity, say) that her person-directed practices are organized around. This allows there to be different persistence conditions for persons across communities without there being confusion about what those persistence conditions are. If you live in a community that organizes their practices around psychological continuity, for instance, you will still be able both to recognize that animalists have different persistence conditions and to identify what those persistence conditions are. You will just believe that they’re wrong to do so, in the same way that the animalist will believe that you’re centering your practices around the wrong property. The conventionalist will look about at both communities tolerantly and say, “What is happening is that different properties need to be preserved for an individual to survive in in these different communities.”[[14]](#footnote-14)

This simple account of survival is made more complex, however, by the fact that the conventionalist also allows for cases of ‘refiguration’ or 'transfiguration' -- cases in which an individual’s person-directed practices shift such that they are now organized around a new property (or set of properties) or relation (or set of relations) in ways that change the persistence conditions for that individual. In certain cases, shifts in one’s person-directed practices can actually generate a new person;[[15]](#footnote-15) following the more recent literature, we will use the term ‘transfiguration’ to describe situations in which the conventionalist takes it that a new person is generated by such a radical shift in person-directed practices.

Before discussing in more detail how transfiguration generates persons, it’s important to take a moment to emphasize the fact that the conventionalist really does mean what she says here. This claim shouldn’t be read metaphorically, as when someone gets a particularly good night of sleep and happily announces, “I’m a whole new person!” The process of transfiguration quite literally generates a new person. Pre-transfiguration, there is one person; post-transfiguration, there is both that original person *and* another person who has been generated by the adoption of person-directed practices that center around a different survival-relevant property or set of properties. Depending on your metaphysics, these persons are either co-located until the ‘survival event’ (such as teleportation) that one survives and the other doesn’t, or these persons share person-stages until the survival event. This is one of the most radical consequences of strong conventionalism: persons are partly constituted by conventions, and a significant enough change in the right conventions can bring it about that there is a new person.

To see more clearly how this process is meant to go, let’s return to the teleporter case we considered above, cashed out for ease of explanation in terms of a four-dimensional account of persistence: perdurantism.[[16]](#footnote-16) According to perdurantism, persisting objects are four-dimensional space-time worms constituted by distinct temporal parts. A person, then, is the mereological fusion of a number of distinct temporal parts -- person ‘stages’ -- each of which is a three-dimensional slice of a person; a person persists by having multiple stages at distinct times.

As noted above, the physicalist and the Lockean communities have distinct survival conditions: by her own lights, a physicalist will not survive transportation, but a Lockean will. Now suppose that a physicalist knows she will be forced to transport in six months, as part of a relocation program implemented by the bureaucratic Lockeans in charge of the government. Knowing this, the physicalist chooses to undergo transfiguration: she changes her person-directed practices over the course of the next six months to accord with those of the Lockean and steps into the transporter believing that she will survive the event.

*Does* she survive? At this point, conventionalists split into two camps, depending on their views about which person’s perspective is the most appropriate from which to judge survival. The first view, which we’ll call ‘weighted conventionalism’, claims that the answer in this situation is an unequivocal ‘yes’, on the grounds that the person-directed practices of the person *at the time she steps into the transporter* are what matter. Mark Johnston (1989), for instance, claims that one should judge whether or not A survives from the perspective of the person at the time she undergoes the ‘survival event’ (e.g., stepping into the teleporter), regardless of what her person-directed practices were centered around at an earlier time. Thus, if a Lockean steps into a teleporter, the same Lockean steps out on the other side, regardless of whether she was a hard-core physicalist in the past. If a person is a physicalist at the time she steps into the teleporter, however, she dies, and it is a new person who steps out on the other side, regardless of her previous views.

The second view, which we’ll call ‘egalitarian conventionalism’, claims that weighted conventionalism falls prey to an Arbitrariness Objection -- namely, that it’s objectionably arbitrary to privilege the perspective of certain person-stages over others -- and that the correct answer to the question of whether A survives is more complicated. (We will discuss the Arbitrariness Objection in more detail in section 5.) Braddon-Mitchell and West (2001), Braddon-Mitchell and Miller (2004), and Miller (2009), for instance, all hold that whether A survives stepping into the transporter depends on whom you ask. If you consider things from the perspective of the **pre**-transfiguration physicalist (or, more precisely, those person-stages that organise their person-directed practices around physical continuity) then the answer is ‘no’. The person with those person-directed practices dies when she enters the transporter. If, however, you consider things from the perspective of the **post**-transfiguration stages that organise their person-directed practices around psychological continuity, then the answer is ‘yes’.[[17]](#footnote-17) The reason for these different answers is that the transfiguration event brought a new person into existence, and these two distinct perspectives correspond to the perspectives of two distinct persons. The person-stages of the new Lockean and the old physicalist overlap from transfiguration until the teletransporter event. At the moment of teleportation, however, the physicalist ceases to exist, while the Lockean survives. *[see figure 1]*

An important consequence of egalitarian conventionalism, and one we will focus on extensively, is that decisions about whether or not one should transfigure to survive some future event become rationally intractable -- that is, that one cannot rationally decide whether to transfigure. Why? Because transfiguration generates a new person that shares stages with the old person until the ‘survival event’ (e.g., teletransportation) occurs. For the time between transfiguration and the survival event, the reference of ‘I’ will be ambiguous between these two persons. In order to make rational decisions, however, we need to know who we are. So it is a precondition of deciding whether or not to transfigure that one knows what ‘I’ refers to at the time *t* at which one is deciding. But what ‘I’ refers to at *t* depends on whether or not you decide to transfigure. There are two ways of understanding the resulting situation: 1) in order to decide whether or not to transfigure you have to have already decided whether or not to transfigure; 2) you’re being asked to make a choice about ‘your’ best interests when there’s no fact of the matter yet about who ‘you’ are. Either way, there is no reasonable way to make the relevant decision.[[18]](#footnote-18)

According to the egalitarian conventionalist, this feature is an advantage of the view: it explains why we have so much difficulty settling on an answer to puzzle cases about personal identity, and why it can seem as though there is no one, correct answer to such cases. (In short, because there’s not!) We believe, however, that the rational intractability of transfiguration decisions is a cost to the view. In particular, there is at least one common case of transfiguration that philosophers generally take to be rationally tractable. It is to this case -- Pascal’s Wager -- that we now turn.

**3. The Wager**

Following Hajek (2003), we take Pascal’s wager to be a three-premise argument of the following form:

1. Rationality requires you to give positive probability to God’s existence.

2. The following decision matrix:

|  |  |  |
| --- | --- | --- |
|  | God exists | God does not exist |
| Wager for God | Infinite gain | Finite loss |
| Wager against God | Finite loss | Finite gain |

and,

3. Rationality requires you to perform the act of maximum expected utility.

The rows in the decision matrix correspond to actions. In this case, the actions are decisions: the decision to believe in God versus the decision not to believe in God. The columns represent outcomes, or states of the world. Expected utility of an action is determined by considering the value (i.e. utility) of each particular outcome multiplied by its probability, and then combining the values for each cell in a row to determine the overall expected utility of the action. The basic idea is that coming to believe in God has a greater expected utility than the alternative. The expected utility of believing in God is, at worst, infinite gain. The expected utility of not believing in God is, at best, finite gain.

The reason the wager is relevant for our discussion is that Pascal holds that, if one chooses to take the wager, one should consistently engage in practices that are likely to transform one’s beliefs about God (e.g. attending Church, reading the Bible, courting mystical experience and so on). A change in religious belief elicited by Pascal’s wager thus appears to constitute a shift in the very kinds of person-directed practices that are targeted by conventionalists.[[19]](#footnote-19) Furthermore, the aim of this action is belief in God *in order to increase one’s chances of survival* and eternal reward after an event one did not previously anticipate surviving (viz, physical death). Thus, if taken seriously, Pascal’s wager requires transfiguration. More importantly, the decision that one must make in deciding whether or not to take the wager is a decision about whether or not to transfigure.

As we discussed in the previous section, however, decisions about whether to shift one’s person-directed practices to survive an event one would not otherwise survive are rationally intractable on egalitarian conventionalism. So it follows that the wager is rationally intractable. If egalitarian conventionalism is true, then, one cannot rationally decide either to accept or to reject Pascal’s wager.

This is a problem, for while there is a great deal of disagreement about the wager, it is widely held that one can, at the very least, rationally decide tochoose one of the options that the wager offers. What philosophers typically disagree about is what decision theory proscribes in this situation, not whether the wager is rationally decidable.

At any rate, while we admit that it is open to the egalitarian conventionalist to accept that Pascal's wager is rationally intractable, we believe that this is a cost to be avoided and thus that it is worth exploring the options open to the conventionalist for doing so. In the following section, we consider three ways this might be done. Each one of them comes at some significant cost to the view, however, and so we argue that the egalitarian conventionalist should avoid making this move.

**4. Looking for Rational Tractability in the Face of Radical Restructuring**

*4.1. Attempt One: Adopt Endurantism instead of Perdurantism*

The rational intractability of transfiguration originates from the sharing of person stages between distinct four-dimensional person worms -- worms that are individuated by their person-directed practices. Initially, then, one might be tempted to solve the problem facing conventionalism by jettisoning perdurantism. For if persons do not *perdure* but rather *endure* by being wholly present at every moment at which they exist, then there will be no sharing of person-stages, and so no referential indeterminacy in decisions involving transfiguration.

The trouble with this move is that any version of endurantism sophisticated enough to model the account of transfiguration we have been considering will give rise to the same sort of referential indeterminacy, even without the problem of person-stage-sharing.

Consider the teleporter case again, this time from an endurantist perspective. Person A is a physicalist at time *t1* who transfigures into a Lockean at time *t3*, and steps into the transporter at time *t6*. At *t6*, person B (who is psychologically but not physically continuous with person A) steps out of the transporter. [*see figure 2*] On egalitarian conventionalism, the original physicalist, A, is correct at *t1* to think that she won’t survive the transporter event at *t6*. But the Lockean is equally correct at *t4* to think that person A survived the transporter event at *t6* -- and, furthermore, that she is identical to person B.

This sort of egalitarianism is difficult to square with endurantism, but it can be done. Endurantists hold that persisting things, such as persons, are numerically the same thing at every time at which they exist, and numerical identity of this kind is not the kind of thing that can be relative to a perspective. If person A is numerically identical to person B, then A and B are identical across all perspectives. The trouble is that the physicalist thinks she dies when she steps into the transporter, and so she denies that person A is identical to person B. The Lockean, however, thinks that person A survives the transporter experience, and so she maintains that person A isidentical to person B.[[20]](#footnote-20)

One way under endurantism to accommodate the egalitarian conventionalist’s claim that the physicalist and the Lockean are both equally correct is to argue that there is something besides person A that exists at t1 and to which person B is identical -- something that *does* survive transportation. Call this thing, person A\*. Person A\* will need to be just like person A, except that she *does* survive transportation. This means that person A\* will need to be co-located with person A, to be qualitatively identical to person A, to have the same thoughts as A and so on. [*see figure 3*]

As it turns out, however, it is the egalitarian claim that the perspective of the **pre**-transfiguration person regarding whether or not they survive is completely on a par with the perspective of the **post**-transfiguration person that is the real source of the referential indeterminacy, not perdurantism. For the existence of person A\* at t1 introduces the same referential indeterminacy into the picture as if there were stage-sharing going on at t1. When the physicalist at t1 is deciding whether or not to transfigure, the reference for ‘I’ will be indeterminate between person A and person A\*. Both are equally good candidates for the reference of the indexical. Person A will claim at t1 that she will not survive transportation, and Person A\* will claim at t4 that she has survived transportation, and there is no non-arbitrary way to break the symmetry -- at least, not without weighting some stages of a person’s existence more heavily than others and thus abandoning egalitarianism. In order to make endurantism compatible with an egalitarian approach to transfiguration, we must reintroduce the rational intractability we were hoping to avoid.

Because endurantism generates the same results as perdurantism regarding the intractability of Pascal’s Wager, we’ll continue to use the perdurantist framework for ease of explanation. There are at least two further ways of attempting to make the wager intractable on egalitarian conventionalism, however: accepting fragmented persons and separating the decision to transfigure from the actual process of transfiguration.

*4.2. Attempt Two: Accept Fragmented Persons*

Egalitarian conventionalists generally hold both that the referent of ‘I’ is ambiguous for the person choosing whether to transfigure (because what ‘I’ refers to at *t2* depends on whether or not that person decides to transfigure) and that the resulting intractability is a positive feature of the view (because it tracks our intuitions about how difficult transfiguration-type decisions are to make). David Braddon-Mitchell and Caroline West (2001) do consider a possibility, however, that appears to eliminate this ambiguity and renders cases like Pascal’s Wager decidable. The general idea is that the person stage making the decision about whether to take the wager should be thought of making the decision *as* and *for* a local aggregate of *individual person stages* and **not** on behalf of the whole person. In this case, there’s no ambiguity of reference for ‘I’ (and hence no intractability) because ‘I’ refers determinately to the ‘fragmented person’ -- the aggregate of person-stages making the decision at *t2 --* and not indeterminately to either of the persons that aggregate is potentially part of.

On a standard four-dimensionalist perdurantist account of persons, the decisions of a given temporal part are made on behalf of the entire four-dimensional worm. So, for example, when one is deciding whether or not to dye one’s hair purple, the local aggregate of person-stages involved in deliberating and making a final choice are considered one part of the four-dimensional worm that’s making the decision. According to the ‘fragmented person’ view, however, the decision at stake in potential transfigurations is importantly different. Rather than being a decision on behalf of an entire four-dimensional worm, the decision is made by the aggregate of person-stages involved in deliberating and making a final choice, *considered as its own ‘fragmented person’*, about which possible person (e.g. the physicalist or the Lockean, the atheist or the theist) that aggregate of person stages wants to be a part of. The relevant analogy here is club membership: when a ‘fragmented person’ makes a transfiguration decision, it is deciding which group of person-stages it wants to belong to. If, for instance, it decides to adopt theistic conventions, then it is deciding to be a part of a four-dimensional worm, the future stages of which all share the relevant conventions.

Because it eliminates the referential indeterminacy of ‘I’ in transfiguring choices, the ‘fragmented person’ view makes Pascal’s wager rationally tractable. Indeed, it seems as though it would be highly rational for the individual person-stage at t2 to choose to be a stage of the 4D worm with the possibility of infinite utility. If you’re going to join a club, after all, why not join the club that yields the promise of infinite utility for its members? If that’s right, then not only is the wager rationally decidable on this view, but we seem to recover the standard result as well.

Although the ‘fragmented person’ view renders the wager tractable, however, it does so at what seems to us for several reasons to be an objectionably high cost. First, accepting this solution pushes the conventionalist heavily towards perdurantism. For while there may be some way to understand this talk of ‘club membership’ using an endurantist framework, we doubt that the account will be straightforward. The trouble is that the fragmented person’s decision about ‘which club to be a part of’ amounts, on endurantism, to deciding which thing one is numerically identical to. For many, this will render the notion of numerical identity worryingly perspectival or relativistic.

Second, as Braddon-Mitchell and West themselves admit, this is “the most revisionary consequence” of their already revisionary view, for it requires us to revise the rules for the use of ‘I’.[[21]](#footnote-21) When contemplating whether to become a Lockean to survive upcoming teleportation, for instance, I would have to think of myself as a ‘fragmented aggregation of person-stages’, rather than as a person. Braddon-Mitchell and West comment, “Jane, the person, cannot ask herself if she survives the transfiguration of all that she cares about, but Jane, the small fragmented group of person-stages, can ask which person she wants to be a part of.” As they go to point out, this has the rather poignant consequence that, “When we are faced with the most difficult decision we are more alone than ever. For we act, not as a person, but as a fragment.”[[22]](#footnote-22)

In the case of Pascal’s wager, finally, this move has particularly unintuitive consequences. Person-stages are quite short-lived, which is precisely why one has to think of oneself on this view as an aggregate of such stages. Suppose that you (qua aggregate of person stages) accept the wager, and that the four-dimensional worm you decide to be a part of ‘wins’ everlasting life and infinite reward. By the time that person begins to experience infinite utility, ‘you’ (qua aggregate of person stages) will no longer exist. In other words, you can decide to join a club where some future members of that club may gain infinite utility -- but you do so knowing that *you* will not be in a position to get this utility. In this light, it appears that the choice facing you (qua aggregate of person-stages) is a choice which can at best yield only *finite* utility. The ‘club’ that you can choose to join has the chance for infinite utility, but you as decision-maker have only the chance for the finite good that comes from being a member of a club with that future possibility.

Pascal’s wager thus rendered boils down to a choice between *finite* utilities. While the wager is rationally decidable on the fragmented person view, then, the view fails both to recover the standard answer to that wager and to preserve the infinite utility option. This seems, to us, to be an unpalatable result of the fragmentary view.

*4.3. Attempt Three: Exchange an (indeterminate) ‘I’ for an (determinate) ‘I’*

There is at least one further way for the egalitarian conventionalist to make Pascal’s wager rationally tractable: she can **a)** accept that the referent of ‘I’ is *usually* indeterminate in cases where someone is deciding whether to transfigure but deny that ‘I’ is *always* indeterminate in such cases, and then **b)** claim that Pascal’s wager is one of the cases in which the reference of ‘I’ is determinate. The rational intractability of Pascal’s wager, recall, is generated by the sharing of person stages (or, assuming endurantism, co-located persons) at the time of the decision about whether to transfigure. But one can hold that transfiguration generates new persons while denying that there are always two persons present at the time of the decision to transfigure. Perhaps in particularly radical cases of transfiguration, the person generated reaches only as far back as the transfiguration *event* -- the actual shift in the conventions that brings a new person into existence -- and not all the way back to the *decision* to change the relevant conventions. In these cases, there would be no stage sharing of persons at the time of decision-making, and so the referent of ‘I’ would not be indeterminate. Decisions about such cases of transfiguration could then be rationally decided in the normal way.

On this version of egalitarian conventionalism, transfiguration would *often* generate persons that reach all the way back to the decision stages -- but it would not *always* do so. What would distinguish the one sort of transfiguration from the other is the extremity of the shift in person-directed practices.

To see this more clearly, it will help to imagine two different transporter cases before returning to the case of the wager. In the first transporting case, Audrey is the physicalist who’s just learned that she’ll be transported off-world in six months as part of the Lockean government’s relocation program. Audrey lives in a fairly ecumenical community, has some Lockean friends, and isn’t too fussed (if still genuinely torn) about the idea of transfiguring to become a Lockean herself. Her person-directed practices do become centered around a new property (namely, psychological continuity) as a result of her decision to transfigure, but this is not an extreme transfiguration. After she becomes a Lockean, and even after she steps out of the transporter, Audrey\* considers her pre-Lockean, physicalist stages to be part of her. In this case, the new person created by the transfiguration event reaches back to the decision stages -- and, in fact, all the way back to what both physicalist and Lockean Audrey consider the beginning of her existence. [*see figure 5*] This is the standard case of transfiguration, where the decision facing Audrey at *t2* is intractable on egalitarian conventionalism.

Now imagine a second transporter case, in which Adam is the physicalist who’s just learned he’ll be relocated via transporter in six months. Adam grew up in a strict physicalist household, in a conservative physicalist town, and he’s been warned for years about the dangers of transporter use and the subversive Lockean government that is touting it as a convenient, cheap means of travel. His family and community would mourn his conversion to Lockeanism as they would a physical death, and they would never consider the person who steps out the other side of the transporter to be Adam, even if he had all the requisite memories. Suppose that, despite all of this, Adam decides to transfigure into a Lockean before he steps into the transporter. (Perhaps he’s just scared enough of death and just uncertain enough about transportation to give it a try.) Now, imagine further that, post-transfiguration, Adam\* is transported to a conservative Lockean community on a distant planet, where he is surrounded by other zealous Lockean converts. Adam\* does not accept any of the pre-transfiguration stages as part of him, in the same way that pre-transfiguration Adam would not consider post-transfiguration stages as part of *him*. In such a situation, the pre-transfiguration stages that exist at the time of deciding are never part of the post-transfiguration person. [*see figure 6*] Adam is the only person who exists at t2 (the decision to transfigure), and so there is no uncertainty about the reference of ‘I’ in this case.

Applied to the case of Pascal’s wager, this version of egalitarian conventionalism could make the wager rationally tractable if it considered religious conversion a case of extreme transfiguration. The change in person-directed practices that accompanies a shift in religious belief would need to so radically alter the original person that even the stages that decided to undergo the relevant transfiguration would not be considered parts of the freshly minted, deeply religious person.

The attractiveness of this view obviously depends on how plausible it is to imagine that the sort of religious transfiguration called for by Pascal’s wager is a case of extreme transfiguration. One reason to think that transfiguration of this kind *is* quite extreme is the thought that a shift in religious beliefs requires a radical restructuring of almost everything that a person believes. Jon Kvanvig, e.g., argues for a view on which religious faith appears to entail radical restructuring. In his words, having faith involves

"an overall metaplan that structures and organizes whatever other plans one might have in terms of some ideal. Religious faith thus aims at the full integration of a life in relation to an all-encompassing ideal...The unification of the specific items of human behavior from one moment to the next in teleological terms counts as an expression of religious faith only when the ideal involved contains in its essence the goal of structuring all plans, purposes, and goals in terms of a single pursuit or ideal."

If accepting Pascal’s wager truly involves restructuring “all plans, purposes, and goals” in pursuit of a new-accepted goal (e.g., living the religious life), then it seems not unreasonable to suppose that a person’s person-directed practices will change in correspondingly radical ways.

Imagine, for example, an existentialist atheist named Sara. Sara lives in an anarchist collective, and her chosen life-project for many years has involved visiting the pediatric oncology ward at her local hospital and attempting to persuade terminally ill children that God does not exist. (She is strongly opposed to anyone’s living in bad faith, and she wants the children she visits to have the chance to give their lives meaning while they still can.) Now, suppose that one of Sara’s roommates challenges her to consider Pascal’s wager and that she agrees to do so, taking the time to study the terms of the wager seriously, and that she eventually accepts the wager and enacts the corresponding shift in person-directed practices. As in the case of the conservative physicalist, Adam, it seems not unreasonable to say that Sara’s conversion to theism would bring about such a radical change to her person-directed conventions that it would count as an extreme transfiguration. If so, Pascal’s wager is rationally decidable for Sara, because the aggregate of person-stages that decided to take the wager will not be parts of the post-transfiguration person. Post-conversion, Sara\* does not consider pre-conversion person-stages to be parts of herself. [*see figure 7*]

There are a few troubles with this suggestion, however. The first is that Sara’s is a rather unusual case. While it might be plausible to think that her transfiguration to become a God-fearing individual is extreme enough to prevent stage-sharing during the decision process, this would seem far less plausible in most other cases. Consider, for instance, Steve. Steve was raised in a Catholic household, sent to Catholic schools, and, for the most part, has adopted many of the moral beliefs and person-directed practices that accompany such a culture. But he does not believe in God. Suppose that Steve is offered Pascal’s wager and, in light of the wager, decides to transfigure and begins actively doing what he can to encourage belief in God.[[23]](#footnote-23) Steve already has many -- perhaps even most -- of the person-directed practices associated with the kind of meta-plan that Kvanvig has in mind. In this case, then, even when he converts, Steve does not seem to undergo the sort of radical transfiguration that would render Pascal’s wager rationally decidable. He will consider his transfiguration decision stages parts of the new person, just as much as the old, and so there will be stage sharing of the kind that renders transfiguration rationally intractable. [*see figure 8*]

Second, it would be strange indeed if, on egalitarian conventionalism, the only cases on which Pascal’s wager turned out to be decidable were cases in which accepting the wager would constitute such a drastic change that the resulting person wasn’t even able to recognize the person who accepted the wager as herself. Pascal’s wager is meant to be the kind of option that you can reasonably consider without worrying that accepting it will mean the end of your existence! Yet, for both Adam and Sara, the shift in conventions is so radical a break from the previous set that the post-transfiguration person and pre-transfiguration person don’t actually share any stages. In these situations, what happens is that Adam and Sara cease to exist, full-stop, at transfiguration. The new persons that are generated have physical continuity and some psychological continuity with the original persons, but they share no stages with the original persons, and cannot be seen as having survived the transfiguration. But that means that Adam’s decision to become a Lockean and Sara’s decision to accept Pascal’s wager lead to their ceasing to exist at the moment they successfully transfigure. This is why accepting Pascal’s wager might be (almost) the last thing you ever do.

Finally, even if transfiguration via religious beliefs is extreme for some individuals (in a way that entails their death!), it is unlikely to be so for most individuals. So the wager will remain intractable for the vast majority of people in a position to contemplate accepting it. Thus, even if the problem posed by the wager for the egalitarian conventionalist is solved for *some* transfiguring persons (at the highest possible cost), it is not resolved across the board -- and it certainly is not resolved in a way that is likely to satisfy either the egalitarian conventionalist or her opponents.

**5. Unconventional Conclusion**

In response to these considerations, the egalitarian conventionalist might just bite the bullet and accept the negative consequences of claiming that transfiguration decisions are rationally intractable. In fact, it’s possible that the egalitarian conventionalist might see this result as an advantage of her view: “We’ve always thought there was something objectionable about the wager that Pascal gave us,” she might say, “And now we know why: the wager is objectionable because -- when we dig into the metaphysics of personal identity -- we discover that the decision to take or leave the wager is rationally intractable!”

We find it implausible to suppose that the solution to Pascal's Wager should hang on facts about personal identity, though. For one thing, if it did, there would be a variety of ways to manipulate responses to the prospective decision. One could, for instance, simply dig in one's heels and claim that resurrection is not person-preserving, and thus that there is no sense in which anyone is in the running to get infinite utility from God. Or one could identify persons with immortal immaterial souls and claim that anyone who didn’t accept the wager was being irrational because infinite utility is always at stake. This is, however, simply not how the discussions of Pascal’s wager go. Even if the wager were to turn on metaphysical facts about persons, moreover, it would be surprising if such facts rendered the wager rationally *intractable*, given that it’s typically portrayed as a standard example of a decidable problem.

Furthermore, even some card-carrying egalitarian conventionalists find this result of their view objectionable. Kristie Miller, for instance, one of the leading proponents of this view, has recently resisted the conclusion that transfiguring decisions are intractable, and proposes two ways of making such choices rationally decidable.[[24]](#footnote-24) The first involves allowing that it’s permissible to weight the utility of some stages of the person more heavily than other (future) stages of the person, particularly in situations where it’s indeterminate whether or not those future stages will be continuer-stages of the person. Miller acknowledges, however that this move involves sacrificing egalitarianism, which she considers a significant cost, and so she also proposes another alternative which she calls ‘prudential relativism,’ which she claims renders the decision tractable relative to a particular perspective.[[25]](#footnote-25) (We do not find Miller’s proposal convincing. For our purposes, however, what’s significant here is simply her attempt to resist the conclusion that transfiguration decisions such as Pascal’s wager are rationally intractable.)

The egalitarian conventionalist could give up her even-handed treatment of person-stage perspectives and retreat to a Johnston-style ‘weighted conventionalism’ (according to which transfiguration is rationally tractable because what counts for survival are only the conventions of the person at the time at which she undergoes transfiguration). As we’ve seen, however, this retreat would require sacrificing one of the central intuitions of the view.

Furthermore, it’s not clear that even weighted conventionalism delivers the correct result for Pascal’s wager. We have argued that religious transfiguration can sometimes have an radical restructuring effect, as in the case of Sara the atheist. In such cases, the restructuring effects of transfiguration may be so extreme as to turn some religious transfigurations into a kind of suicide: the person who comes out of the transfiguration is metaphysically (not just metaphorically) a different person from the one who went in.

Of course, not all religious transfigurations will be of this kind, as in the case of Steve. Pascal’s wager, then, raises an important general question for conventionalist views. For we are not always in a position to know whether the kind of religious transfiguration involved in our coming to believe in God will be so extreme as to be a kind of suicide. So, on the conventionalist view, we may not be in a position to know whether we would survive the wager or not. This, in itself, may make the wager rationally undecidable, even without indeterminacy about the reference of ‘I’.

*Christina Van Dyke (Calvin College)*

*Sam Baron (University of Sydney/University of Western Australia)*

**Bibliography**

Braddon-Mitchell, David and Miller, Kristie. (2004). ‘How to be a Conventional Person’ *The Monist* 87(4): 457-474.

Braddon-Mitchell, David and West, Caroline. (2001). “Temporal Phase Pluralism.” *Philosophy and Phenomenological Research* 62(1): pp 1-25**.**

Eklund, Matti. (2004). “Personal Identity, Concerns, and Indeterminacy.” *The Monist* 87 (4):489-511.

Hajek, Alan. (2003). “Waging War on Pascal’s Wager.” *Philosophical Review* 112: 27-56.

Hume, David. (1978) *A Treatise of Human Natur*e (eds.) L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press.

Johnston, Mark. (1989). “Relativism and the Self” in Krausz, Michael (ed.), *Relativism: Interpretation and Confrontation*. Notre Dame: University of Notre Dame Press, pp. 441-72.

Kvanvig, Jonathan. (forthcoming) “Affective Theism and People of Faith.”

Lewis, David. (1976). “Survival and Identity” in Rorty, A. (ed.), *The Identities of Persons*. Berkeley: University of California Press, pp. 17-40.

Merricks, Trenton. (2001). “Realism about Personal Identity over Time.” *Noûs* 35: supplemental volume 15 (*Philosophical Perspectives*, 15, *Metaphysics* edited by James E. Tomberlin). 173-187.

Miller, Kristie. (2009). “Deciphering Personal-identity Conventionalism” Special Issue *Rivista di Estetica “Convenzioni”* edited by Elena Casetta and Achille Varzi 44:2 59-85.

Miller, Kristie. (forthcoming). “Prudence and Person Stages.” Special Issue *Inquiry*, Choice Over Time: Diachronic Practical Rationality, edited by Sergio Tennenbaum

Olson, Eric. (1997b). “Relativism and Persistence.” *Philosophical Studies* 88:141-162.

Olson, Eric. (2007). *What Are We?* Oxford: Oxford University Press.

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

Unger, Peter. (1990). *Identity, Consciousness, and Value*. Oxford: Oxford University Press.

van Inwagen, Peter. (1998). *The Possibility of Resurrection and Other Essays in Christian Apologetics*. Boulder, CO: Westview Press.

Zimmerman, Dean. (1999). “The Compatibility of Materialism and Survival: the ‘Falling Elevator’ Model.” *Faith and Philosophy* 16: 194-212.

1. The standard animalist answer to the metaphysical question of what persons is that “each of us is numerically identical with an animal; there is a certain organism, and you and it are one and the same” (Olson 2007, 24). The answer it provides to the question of what the necessary and/or sufficient conditions for our persistence might be is that we persist so long as the animal to whom we are identical persists. [↑](#footnote-ref-1)
2. One way to bring animalism and conventionalism into conversation is to treat them as (respectively) realist and antirealist accounts of personal identity and then to argue for one general sort of account over the other. This is, e.g., roughly the strategy Trenton Merricks adopts in “Realism about Personal Identity over Time” (2001). The difficulty with this strategy is that it more often brings out a clash of intuitions about what a plausible account of personal identity must look like (to the tune of “That’s not an objection; that’s the view!”) than gets at the heart of the philosophical plausibility of the competing views. [↑](#footnote-ref-2)
3. The main argument in *favor* of animalism is the solution it offers to what’s often referred to as the “too-many thinkers” problem. The problem, briefly, is this. It is evident that human animals exist. Moreover, it is equally evident that such animals are capable of conscious experiences, thought and so on. But we typically think that persons are also capable of conscious experience and thought. We also believe that persons are (at least) located wherever there are human animals. If persons are not identified with human animals, however, then we are forced to duplicate thinkers: on the one hand, we have the human animal thinking its thoughts at some region of space-time *R* and, on the other hand, located exactly where the human animal is (i.e. in the region *R*), we have a numerically distinct person thinking its thoughts. In short, if animalism is false then every one of us appears to be sharing the same region of space-time with a thinking animal. If animalism is true, then no such proliferation of thinkers is necessary: persons are identified with human animals, and so there is only ever one thinker located in *R*. Not so if animalism is false. See [Olson 1997--exact pps?] for a detailed discussion. [↑](#footnote-ref-3)
4. Indeed, pressure to account for post-mortem survival has famously led to outrageous-sounding accounts like van Inwagen’s “Body-Snatching God” (1998), where God personally ensures our survival by removing our bodies (or whatever part of our bodies is essential to our existence) the moment prior to our death and replacing them with qualitatively indiscernible duplicates and Zimmerman’s “Falling Elevator” model (1999), where God causes us to ‘fiss’ at the moment prior to death in such a way that the immanent causal relations between us and the resurrected person guarantee that the resurrected person will at least be our closest continuer. [↑](#footnote-ref-4)
5. Pg. 261 of the Selby-Bigge and Nidditch edition (Oxford University Press, 1978). [↑](#footnote-ref-5)
6. La la la references. [↑](#footnote-ref-6)
7. Kristie Miller (2009) distinguishes between four distinct versions of conventionalism: multiple candidate semanticism, multiple candidate conventionalism, strong conventionalism and eliminative conventionalism. Perhaps the most popular version of conventionalism (and the one that provides the strongest solution to jumpy cases of personal identity) is the third of these: strong conventionalism, and so it is this version of conventionalism that we intend to focus on here. [↑](#footnote-ref-7)
8. We’re following the standard conventionalist line here in including a body in the mix. Hard core dualists in the Cartesian/Platonic tradition would, of course, deny that persons need to have bodies. On strong conventionalism, this view is coherent if in fact dualists and their communities focus their person-directed practices around immaterial souls and not physical bodies. [↑](#footnote-ref-8)
9. Examples of mere convention in the Lewisian sense include driving on the left (or right) side of the road and marking ‘north’ on the top of a compass or map. See Lewis (1976). [↑](#footnote-ref-9)
10. Conventionalists themselves don’t have much to say (to this point) about what the exact nature of these hard-wired conventions are, but they do mention that they’re the sort of proto-practices that we share with non-human animals, such as basic fight or flight responses:

Some practices such as self regard for future continuers in some basic way may be the sort of deeply fundamental practice also exhibited by non-persons or proto-persons, and be hard-wired into the brain. They won’t be malleable, nor will they be a matter of indifference. Other practices such as having self-regard for a continuer with which one is psychologically but not physically continuous, seem not to be hard-wired in this way and thus malleable. (Braddon-Mitchell and Miller 2004, p. 133) [↑](#footnote-ref-10)
11. See Braddon-Mitchell and Miller (2004) for this sort of response to many/most of the objections against conventionalism raised in Merricks (2001). [↑](#footnote-ref-11)
12. “Prudence and Person Stages,” forthcoming in *Inquiry*. [↑](#footnote-ref-12)
13. Braddon-Mitchell and West, pg. 61. As they go on to note, “When enough of these roles are played by some property, then that property is what it takes to survive for that person or community.” [↑](#footnote-ref-13)
14. This, obviously, raises the rather pressing point of the relation between personal and social conventions in determining both when a person exists and through which conditions a person persists. Much to our own frustration, space considerations force us here to follow the current literature in passing over this topic and focusing this paper on a simplified, ‘toy’ version of this theory. [↑](#footnote-ref-14)
15. When this occurs is a matter of debate among conventionalists. Johnston (1989) uses the word ‘refiguration’ to indicate that it is usually the same person who considers, adopts, and then enjoys the results of the more permissive survival conventions, allowing that the process generates a new person only in particularly radical and unusual situations. Braddon-Mitchell and West (2001), Braddon-Mitchell and Miller (2004), and Miller (2009), however, all use the term ‘transfiguration’ to mark their belief that such a shift in conventions always generates a new person. [↑](#footnote-ref-15)
16. As we discuss in section 4.2, it is possible but more complicated to model cases of person-generating transfiguration in a non-perdurantist metaphysics. [↑](#footnote-ref-16)
17. The post-transfiguration stages might not consider all the pre-transfiguration stages parts of the same person. In the situation where the Lockean doesn’t consider the pre-transfiguration stages part of *her* 4-D ‘worm’, she’ll see herself as having come into existence when the transfiguration happened. Similarly, if teleportation somehow produces a person with strongly physicalist conventions, such a person would consider herself as having come into existence when she stepped out of the teleporter, since she would not be physically continuous with the person who had stepped into the teleporter. [↑](#footnote-ref-17)
18. As Braddon-Mitchell and West put it, “The standard account says (roughly) that ‘I’ refers to the person who uses it. Translated into four-dimensional talk, this amounts to ‘I’ refers to the person of whom the person stage who utters it is a part. But in cases of transfiguration, there either is more than one such person; or else, which person(s) the stage is a part of will be a consequence of this decision, but the answer must be had prior to the decision” (80). [↑](#footnote-ref-18)
19. Al Hajek, e.g., specifically characterizes the ongoing action required by the wager as “adopting a certain set of practices and living the kind of life that fosters belief in God” (28). [↑](#footnote-ref-19)
20. In the case of the wager, the situation looks like this: person A is an atheist at t1 who transfigures into a theist at t2 and dies at t3. At t4, person B -- who is psychologically (and possibly even physically) continuous with person A -- is resurrected and begins experiencing infinite utility. On conventionalism, the atheist denies that person A is identical with person B, on the grounds that person A ceases to exist at t3 (the event of physical death); the theist claims that person B is identical with person A, who survives death via God’s resurrecting powers. [↑](#footnote-ref-20)
21. In their words, “The revision would be something like ‘I’ refers to the continuing person who utters it, or in the case of intractability, some aggregation of person-stages around the utterer. So, in some cases, ‘I’ picks out not a person at all, but an aggregation of stages” (80). [↑](#footnote-ref-21)
22. “Temporal Phase Pluralism” pp. 82-3. [↑](#footnote-ref-22)
23. Steve can’t, of course, *force* himself to believe in God. As we saw above, the original wager involves adopting person-directed practices that encourage and/or support such belief; just adopting those practices will not be enough to count as transfiguration, however. Steve must begin to believe in God (and perhaps even the Christian doctrine of the resurrection of the dead) in order to count as having transfigured. This case is not as dissimilar from the cases of the physicalist transfiguring into the Lockean as might at first appear. The physicalist can change her person-directed practices to a great extent, but in order to survive stepping into the transporter, she must believe that psychological continuity is what matters for her survival. See [insert references] for further discussion of the extent to which belief is voluntary. [↑](#footnote-ref-23)
24. “Prudence and Person Stages” (forthcoming). [↑](#footnote-ref-24)
25. Prudential relativism is meant to preserve the egalitarian intuition while also making transfiguration decisions tractable. In short, the idea is that although decisions like Pascal’s wager are intractable from a stage-independent stance, they will be tractable from a the perspective of a particular stage. In Miller’s words: “According to prudential relativism it is impossible to offer a single, correct, answer to the question: should person-stage, P, ϕ at t? For according to prudential relativism there is no stage-independent stance from which to evaluate whether a person-stage ought to ϕ. Any such evaluation is always relative to the conventions deployed by some stage or other. Since, according to conventionalists, there are multiple equally good conventions for a person-stage to deploy, it follows that, with respect to at least some claims of the form ‘P ought to ϕ’, said claims will be true relative to the conventions of one person-stage, and false relative to another...That does not make the decision of whether to ϕ intractable: if P is making the decision, then, plausibly, P will (rationally) decide to ϕ just in case relative to P’s standards, P ought to ϕ. It does, however, make it impossible to say whether, independent of any particular person-stage’s conventions, P ought to ϕ.” Although this is an intriguing suggestion, it’s not clear that it solves the original problem, insofar as it doesn’t give the person actually struggling with the decision a clear way to resolve her question. We take it that the person was always able to see that, if she were to think of the potential transfiguration from the perspective of Audrey, say, vs. Audrey\*, she would be able to make a decision.What matters is which perspective she should adopt for the purposes of making the relevant decision. [↑](#footnote-ref-25)