TRANSFORMED BY FAITH

Rebecca Chan

Appealing to self-interest is a common way of justifying the rationality of religious faith. For instance, Pascal's wager relies upon the expected value of choosing the life of faith being infinite. Similarly, many contemporary arguments for the rationality of faith turn on whether it is better for an agent to have faith rather than lack it. In this paper, I argue, contra Pascal, that considerations of self-interest do not make choosing faith rational because they fail to take into account the way the self is transformed by faith.

Introduction

Imagine Blaise, who is approached by a powerful deity who makes him the following offer:

> I can make you, Blaise, the happiest man in the world. In fact, I can guarantee that your happiness will be unsurpassable. The catch is that you will become radically different from who you are now. Your life goals, core preferences, and the way you see the world will be different. But rest assured, once you change, you won't mind parting with all the things with which you most strongly, currently identify. Becoming the happiest man in the world will radically transform you.

From the standpoint of self-interest, is it rational for Blaise to prefer that the deity transform him?

Pascal famously held that accepting an analogous offer is rationally required. One ought to wager on God, he argued, *because* the expected value of the wager, for the wagerer, is infinite. Though convincing oneself to believe in God might involve radical change to one's life, it is ultimately in one's self-interest to (attempt to) believe because the expected value of doing so is greater than the expected value of the alternative. For Pascal, adopting a standard decision procedure—one that takes into account the values of outcomes and likelihood of those outcomes in calculating expected values[1]—fully explains the rationality of faith, at least practically

---

[1]I take "standard decision procedures" to include procedures such as "maximize expected value," dominance, and Rawl's maximin. I also take it that maximizing expected value is the most common of these procedures, and is the one that is most relevant to the Pascalian

speaking.[2] For ease of reference, let's call this strategy that makes an act or preference practically rational in virtue of self-interested value the "Pascalian model." Like Pascal, contemporary thinkers like Buchak[3] and McKaughan[4] have argued that faith is practically rational. Though their accounts don't adopt a purely Pascalian model, they are closely related insofar as they incorporate value and practical rationality (and ultimately, I'll have something to say about these accounts, too).

My target in this paper is the Pascalian model—at least as it applies to adopting a life of faith. The core of my argument centers upon whether the link between self-interested value and rational preference holds when agents are confronting decisions involving outcomes in which they will undergo radical changes to their core preferences, life goals, and way they experience the world. These outcomes involve what I call *self-transformation*, and I begin by unpacking exactly what self-transformation is (§1). I then introduce a case about past preferences in which it appears, somewhat paradoxically, that it's rational for an agent to prefer an outcome in which the agent experiences less value than the agent would in an alternative outcome (§2). I then consider various diagnoses of the case in §2, ultimately favoring one that makes use of the concept of self-transformation (§3). On this diagnosis, the Pascalian model is viable only if modified to account for self-transformation. But once modified, it does not require agents to prefer outcomes involving self-transformation. Next, I extend what I say about preferences about the past to preferences about the future (§4). From here, I return to the case of religious faith and suggest that choosing to adopt the life of faith is an instance of choosing to undergo self-transformation. If the Pascalian model does not require preferring outcomes involving self-transformation, then, *a fortiori*, it doesn't rationally require agents to prefer adopting a life of faith (§5). This consequence has interesting implications, which I highlight in the final section of this paper (§6).

Before getting underway, a few preliminaries are in order. First, this paper is concerned with practical rationality and self-interest. This concern is in keeping with the accounts of Pascal and others who raise the question of whether adopting the life of faith is practically rational. For Pascal, it appears that considerations of self-interest determine whether an agent's choice is practically rational. I'm not going to argue for this particular interpretation of Pascal (though I think it's correct), and I'm just

---

view, which this paper discusses. I'll thus proceed with maximizing expected value in mind, though what I say applies to the other views as well. (As we'll see, I challenge whether the values in question connect with the self-interest of the deliberating agent, and this point applies to the other standard decision procedures.)

[2]There is some debate about whether value considerations might also make such belief *epistemically* rational. This paper takes no position on that debate. It is concerned only with practical rationality.

[3]E.g., Buchak, "Can it Be Rational to Have Faith?" and "Faith and Steadfastness."

[4]E.g., McKaughan, "Authentic Faith and Acknowledged Risk" and "Action-Centered Faith, Doubt, and Rationality."

stipulating that the Pascalian model takes considerations of self-interest to be sufficient for determining whether a choice or preference is practically rational. (So if this paper is successful, it shows that self-interest doesn't require agents to prefer the life of faith.) There's an open general question about whether other considerations (such as moral ones) factor into practical rationality. My argument is silent about the scenario where morality factors into practical rationality and makes it the case that agents are rationally required to prefer a life of faith.

Relatedly, this paper is silent on whether it is epistemically rational for agents to adopt the life of faith. Indeed, one major upshot of focusing on practical rationality is that agents can engage in practical reasoning regardless of their other background commitments. This is especially advantageous in the religious context where theists and non-theists often have such intractable differences in these background commitments, such as the commitment to God's existence. As Pascal noted, focusing on practical rationality allows one to sidestep this issue since epistemic uncertainty is typically accounted for (e.g., Pascal's wager only requires assigning a non-zero probability to God's existence). Thus, the Pascalian model (and my critique of it) has broad appeal and applies to theists and non-theists alike.

Second, this paper, especially §3, discusses rational *preferences*. It does so in part because some of the examples considered in this paper involve outcomes that are not strictly speaking live—the agent comparing outcomes may not be able to choose a particular outcome even if that outcome is preferable to competing ones. Talking about rational preferences rather than rational choices may even be more perspicuous since what one prefers is in some sense more fundamental than what one chooses—an agent's choices are practically rational only if they line up with the agent's preferences. In any case, this paper operates under the assumption that rational preferences and rational choice are aligned. If an agent's preference for one outcome over another is rational, it is also rational for that agent to choose that outcome over the other (or would be rational for the agent to choose that outcome if the outcome were live), and vice versa.

Finally, for simplicity, we'll assume that agents are in, and believe they are in, epistemically transparent situations with respect to value and outcomes. So, for instance, Blaise knows that the deity can and will carry through with the promise should Blaise decide to accept the offer. If Blaise accepts the offer, the outcome in which he is transformed and unsurpassably happy will obtain. If Blaise declines the offer, the outcome in which he is untransformed will obtain. Furthermore, Blaise also knows that transformed, unsurpassably happy future Blaise experiences more value than untransformed future Blaise. Something similar holds for other agents in the other cases we'll consider.

Now that these preliminaries are out of the way, let's move on to the question of whether, from the standpoint of self-interest, it is rational for Blaise, or any other agent, to prefer to undergo radical change based on the Pascalian model.

§1 Self-transformation and Practical Identity

1.1 Transformation

Agents sometimes face decisions in which at least one of the potential outcomes involves radical change. Blaise, from our story above, is not alone in facing this type of decision. Potential parents are in a similar situation according to Paul.[5] Becoming a parent is radically transformative, and part of what makes the transition to parenthood so radical is that it involves what Paul calls *personal transformation*—the potential parent's core preferences, life goals, and even the way she perceives the world change.[6] These preference changes aren't merely superficial ones, such as the casual moviegoer's shift in preference from romantic comedies to epic dramas. They occur at a deeper level, like how the self-absorbed potential parent who detests children may find herself dedicating the rest of her life to her child's welfare. When personal transformation occurs, the preference changes that occur are deep, as are the changes to life goals and the way one sees the world. They are similar to Kuhnian paradigm shifts—the point of view of the personally transformed self can't be understood by the pre-transformed self in a first-personal way because the pre-transformed self lacks the framework with which to understand the future self. Because the potential parent would undergo such radical change to the fundamental ways she interacts with the world, the potential parent may not be able to project future parenthood as part of *her* life.[7]

At best, attempting to imagine a future of parenthood would be more like imagining what it would be like for one's current self to be a parent as opposed to imagining what it's like from the perspective of the future

---

[5]For more on the transformative nature of becoming a parent, see Paul, "What You Can't Expect When You're Expecting" and *Transformative Experience.* Other examples abound. In fiction, we encounter characters confronted with decisions such as whether to become vampires or pass on into the afterlife; they wrestle with these decisions without the benefit of knowing exactly how they will transform (though the changes are highly likely to be personally transformative) and what that future will be like for them. Real life also contains decisions that philosophers allege involve transformative experiences: committing to marriage (Paul, *Transformative Experience*), gaining a new sense modality (Paul, *Transformative Experience* and Harman, "Transformative Experience and Reliance on Moral Testimony"), undergoing a gender transition (McKinnon, "Trans*formative Experience"), becoming a member of a different social group (Barnes, "Social Identities and Transformative Experiences"), undergoing religious conversion or apostasy (Chan, "Religious Experience, Voluntarist Reasons, and the Transformative Experience Puzzle"), and making a life-changing choice (Chang, "Transformative Choices").

[6]According to Paul ("What You Can't Expect When You're Expecting"), transformative experiences have two components: epistemic transformation and personal transformation. Epistemic transformation is the discovery of what the experience is like. In this paper, I'm concerned only with the personally transformative component.

[7]To be clear, on Paul's account agents cannot even project what it will be like to be transformed. If Paul is right, then an agent won't be able to imagine herself as a future parent since she can't imagine what it would be like at all. This paper is neutral on whether agents can project what futures in which they radically transform will be like. But even in the scenario where they are, the problem the paper is addressing arises.

self. In more recent work, Paul[8] suggests that there are at least two ways to empathize with our future selves. *Affective* empathy involves projecting our current selves into potential future scenarios. This type of empathy is akin to what we do when we empathize with others. *Cognitive* empathy is richer, and involves experiencing the potential future from the perspective of the future, rather than current, self. But this type of empathy can take place only when the future self is closely related to the current self. The distinction between these two types of empathies roughly tracks a distinction that Velleman[9] draws between what it is like to visualize what Napoleon saw when looking out at Austerlitz and imagining that one is Napoleon looking out at Austerlitz. The latter removes all first-personal traces of the imaginer. Significantly for both Paul and Velleman, we, as our current selves, can't imagine being a future self who has undergone personal transformation from the perspective of the future self. The perspective of the future self is inaccessible because it is not, in a literal way, the present self's own.

Because the potential future self is literally changed beyond recognition from the perspective of the present self, I'll use *self-transformation* to refer to these radical changes. In light of what's been said, we're now in a position to take a first pass at characterizing self-transformation. An agent undergoes self-transformation from t to t' when and only when the self at t cannot "first-personally" identify with the self at t'. This is just a rough characterization we'll refine later, and the first-personal identification condition is meant to capture the inability of the self at t to project the life at t' as their own. Thinking of the lack of first-personal identification in terms of a paradigm shift, a lack of cognitive empathy, or Velleman's sense of inaccessibility are all compatible with this characterization. Changes that are radical enough to cause these phenomena are changes that involve self-transformation.

1.2 Practical Identity and Self-transformation

As radical as changes like parenthood are, these changes are not typically taken to be changes that call into question the continuity of personal identity. There are no extreme physical changes such as body swapping, brain transplants, fission, or teletransportation. Though the agent's psychology may change with respect to preferences and worldview, there's no memory loss or gappy consciousness. Souls haven't been invoked. Even features one might take to be metaphysically essential to the agent's kind, such as rationality, persist throughout the change. This is the case even if one holds a view of essences on which essential features define what that agent's kind is or a view of essences on which essential features are ones that the kind must have in order to exist. *Metaphysically* speaking,

---

[8]See Paul, "The Subjectively Enduring Self."

[9]See Velleman, "Self to Self."

the changes in question aren't existential ones—they don't threaten the existence of the person, metaphysically speaking.

But *practically* speaking, these changes are existentially threatening. Consider Parfit's young Russian socialist who happens to be the descendant of aristocrats and stands to inherit a fortune. He realizes that he is likely to lose his socialist ideals when he becomes wealthy. As he considers what his future self sans socialist ideals will be like, he realizes that he doesn't identify with that future self at all. His instructions to his wife—"if I lose the ideals, I want you to think that I cease to exist"[10]—are telling.[11] Like Paul's potential parent, the young Russian cannot project future aristocracy as part of *his* life. The socialist ideals aren't merely something he believes; they mediate the way in which he fundamentally understands the world and his role in it. Thus, he can't empathize at all with a future self who lacks these ideals that shape his present self. That future self is so foreign to him that it may as well be a completely distinct person.

Agents are shaped by their preferences, life goals, approaches to decisionmaking, and way they see the world, amongst other things. We can imagine an agent who has all of the features that shape their agential life wiped clean by a mad scientist and then replaced with contrary features. For instance, imagine that a mad scientist has taken Parfit's young Russian and wiped away his socialist ideals, love for his wife, drive to further the welfare of those around him, and other features that shape his agential life. These features are then replaced with selfishness, aristocratic indifference towards those around him, and the like. In this scenario, the agential life of the young Russian has ceased to exist, though perhaps the young Russian continues to exist metaphysically.

This distinction between metaphysical persons and practical agents is a significant but often underappreciated one. Perhaps the most well-known articulation of this view comes from Parfit.[12] According to Parfit, facts about personal identity (in the metaphysical sense) are sometimes opaque to us. In hard cases such as fission, we don't know which of the resulting people, if any, are identical to the pre-fission person. Perhaps there isn't even a fact of the matter regarding whether personal identity holds between the pre-fission person and the post-fission people. However, personal identity also *doesn't matter*. What does matter are the facts about the practical agents, and those are more readily available to us than facts about personal identity. Schechtman[13] has advanced a similar distinction between the metaphysical and the practical. For her, there are metaphysical questions about *what* persons are and practical questions

---

[10]Parfit, *Reasons and Persons*, 327.

[11]Parfit primarily uses this case to consider what one ought to do when future preferences conflict with present ones. But it also nicely illustrates the alienation that can occur when an agent undergoes radical changes and loses features that he or she regards as essential.

[12]See Parfit, *Reasons and Persons*.

[13]See Schechtman, *The Constitution of Selves*.

about *who* they are. Questions about the former often center upon the necessary and sufficient conditions for personal identity. But questions about the latter focus on what characteristics are "truly those of the person."[14] These are the characteristics that most centrally make and express who a person is. So, for instance, having socialist ideals is a characteristic that centrally makes and expresses who the young Russian is, though it is not metaphysically essential. Having these ideals is an essential feature of *who* the young Russian is, but not *what* he is.

I hope that what I've said here is enough to make the reader sympathetic to the distinction between what a person is metaphysically speaking and who an agent is practically speaking. Obviously there is much more that can be said, but I won't offer a fully worked out metaphysics of practical agents here. Instead, I'll ask the reader to reflect upon the features that are central to who they are as practical agents. Perhaps unlike the young Russian and his socialist ideals, there is no single feature that can be identified. But if one continues to strip away features, there will be a point at which losing the sum of those features destroys the practical agent. Those feature(s) are *practically essential*. They define who the agent is practically speaking, and the agent, at least as conceived of practically, cannot survive the loss of them.[15] With this notion of who an agent is practically speaking in mind, let us revisit our initial characterization of self-transformation and precisify it into a definition:

> Self-transformation is the change undergone by an agent, S, between times t and t′ when and only when S at t′ lacks some feature (or cluster of features) that S at t regards as practically essential.

Losing these features is sufficient for the S at t failing to first-personally identify with S at t′.[16]

---

[14]Schechtman, *The Constitution of Selves*, 73.

[15]Practically essential features are to practical agents as metaphysically essential properties are to metaphysical persons. For our purposes, it doesn't matter whether essentialness is analyzed modally (roughly, a property F is essential to X iff if X exists, X is F) or definitionally (roughly, a property F is essential to X iff F is part of X's real definition). On either analysis, an agent's losing a practically essential feature (like the young Russian losing his socialist ideals) is sufficient for the agent ceasing to exist, practically speaking. That's enough to get the rest of the paper going.

How a person's practically essential features and metaphysically essential properties relate to each other is another interesting question that I won't attempt to answer for the purposes of this paper. Perhaps there is no connection between them. Alternatively, perhaps all of the necessary properties of a (metaphysical person) are also necessary properties of the associated practical agent. This is easiest to see when essentialness is analyzed modally: if practically essential features are analyzed modally (so F is practically essential to X iff if X (practically) exists, X is F), all of an agent's necessary properties (including the non-practical ones) count as practically essential. In either case, what's important for our purposes is that persons need both their metaphysically essential properties and practically essential features to exist practically.

[16]Losing those features might also be necessary for S at t failing to first-personally identify with S at t,′ but I won't argue for that since the sufficiency claim is all that's needed for present purposes.

A final feature of self-transformation deserves comment. The feature(s) S loses are ones that S at t *regards as practically essential*. S's regarding the feature(s) as practically essential is important for three reasons. First, whether a particular feature—such as having socialist ideals—counts as practically essential varies from agent to agent. Socialist ideals are essential to Parfit's young Russian, but it's not hard to imagine a Bernie bro who holds socialist ideals, but non-essentially. Second, we're in a context where an agent's ability to first-personally identify with a future self is the factor that determines whether the changes the future self undergoes would involve self-transformation. This factor consists only of S being able to imagine what it would be like to lose that feature. If S can't do that—in other words, if having that feature is essential to S's sense of self—then losing that feature is transformative in the sense that we're interested in.

Finally, one might worry that the definition I've offered eliminates the possibility of an agent's being mistaken about their practically essential features. One way to alleviate the concern is to stipulate that the agents in all the cases discussed in this paper are not mistaken. But that's not necessary because what's at issue in this paper is practical rationality. And what counts as practically rational depends, in large part, on the beliefs of an agent. Agents who prefer outcomes that they believe to be the worst possible outcomes for them are practically irrational even if it turns out that the outcome they prefer is better than competing ones. So suppose that there are practically essential features, but that agents can be mistaken about which of their features are practically essential. (In other words, substitute "is practically essential for S at t" for "S at t regards as practically essential.") It's still the case that what matters for evaluating whether agents are practically rational are going to be their beliefs about, or what they regard as, their practically essential features. Thus, even if agents can be mistaken, it is appropriate to focus upon the features agents take to be essential rather than ones that are essential in some objective sense.

Now that we're equipped with an understanding of what self-transformation is, let's turn to why they pose a difficulty for the Pascalian model.

§2 Trouble for the Pascalian Model?

Reasoning on the Pascalian model works well in ordinary cases. Take, for instance, the rather mundane decision I make each morning to drink coffee, tea, or neither for breakfast. If the state of affairs in which I drink coffee for breakfast is more valuable than the state in which I drink tea, then it's in my self-interest to have coffee rather than tea. Furthermore, later in the day when I reflect back upon what I drank during breakfast, it makes sense for me to prefer that I drank coffee to some alternative given that the coffee state was in fact that best state. As far as practical rationality is concerned, it is rational for me to prefer coffee to tea, and to ultimately choose coffee rather than tea. Considerations of self-interest are what make the preference for coffee rational, and the Pascalian model works—both when the preference concerns the past and the future.

But while the Pascalian model does well in ordinary cases, it doesn't do well in cases where the agent would be radically different in the potential outcomes. The model breaks down (or so I'll argue) when the decision point involves self-transformation. To motivate this position, I offer a pair of cases about preferences regarding the past. (I'll extend what I say about these cases to the future in §4.) The first is a variation on the non-identity problem:

> *Non-identity.* Wilma wants to have a child. The doctor tells her that she can conceive now or take a pill and wait a month. If she waits, the child will have LeBron's physical gifts, Parfit's philosophical ingenuity, and Mother Theresa's kindness. If she conceives now, the child will be average with respect to physical and philosophical abilities, as well as altruistic inclinations. Wilma conceives now, and Wilma's child considers whether it would be better for her had her mother waited.

Suppose that, as Kripke suggests,[17] one could not have originated from the union of gametes other than the gametes from which one actually originated such that Wilma's waiting a month would result in a metaphysically different child being conceived. Even if we stipulate that the world would be better from an objective standpoint had Wilma waited to conceive and had a different child a month later, it would be irrational for Wilma's child to wish, for self-interested reasons, that her mother had waited. After all, had her mother waited, she would not exist at all! As long as Wilma's child positively values her life, it is rational from the standpoint of self-interest for her to prefer the world in which she exists to the one in which she does not.

This case demonstrates two things. First, it shows that self-interest links up with a particular kind of value. The overall value of the world, objectively considered, doesn't matter for self-interest. The value in a particular agent experiencing value is not what matters for self-interest either. Wilma's child can recognize that there's value in the experiences of value for the extraordinarily gifted child that would have been born had Wilma's mother waited, but that value is unconnected to what is in the self-interest of Wilma's child.[18] Rather it is the value *for* Wilma's child, or what we've been calling self-interested value, that makes the child's preference that her mother not wait to conceive rational.[19] Self-interested value

---

[17]Kripke, *Naming and Necessity*, 112–113.

[18]These first two types of value that are not self-interested value connect with the what are often labeled as agent-neutral and agent-relative, respectively (see, e.g., Thomas Nagel, *The View from Nowhere*, 152).

[19]Many other sorts of examples capture this point. For instance, from an objective standpoint, it would be better overall for me to donate my organs to five patients who would otherwise die. Moreover, it may be that the value of those five patients getting to experience more value is significant—far more significant than the value of my continued experiences of value. Nonetheless, it would obviously not be in *my* self-interest to make this sacrifice.

is the relevant type of value when considering what an agent is rationally required to prefer.[20]

Second, the case demonstrates that an agent's existence in an outcome is a necessary condition for the obtaining of that outcome being in that agent's self-interest.[21] This point is so obvious that it often goes unstated. But reflection upon how the Pascalian model and other decision procedures work reveals that they implicitly restrict the outcomes under consideration to ones in which the agent exists. The *self* matters in discussions about self-interest. If an agent does not exist in a potential outcome, that outcome obtaining is not going to be in the self-interest of that agent. So for instance, in the case of Wilma's child, the child can set aside outcomes in which she does not exist, such as the one in which her mother waits to conceive.

However, metaphysical identity plus self-interested value fails to capture the entirety of what we care about when it comes to our rational preferences. Here's a second case that's aimed at showing that we also care about who we are, practically speaking:

> *Past Trauma.* Jon suffered childhood trauma. He eventually overcame those experiences such that those experiences shape who Jon has become. We can imagine an alternative past in which Jon's early life went better and did not contain that trauma. Further, we can imagine that the better past does not detract from the value of the later stages of his life. Thus, non-traumatized Jon would have a life that contains more value in the past and overall, though he would also be a very different person from actual Jon. Jon now considers whether he prefers that his

---

[20]In the transformative experience literature, there's an additional wrinkle with respect to the value relevant for self-interest. According to Paul (*Transformative Experience*), the relevant values must be assigned by an agent based on what the experience will be like for that agent. These assigned values are "subjective values." In cases of self-transformation, present agents cannot assign subjective values to the experiences of the future transformed agent since the perspective of the future agent is inaccessible in virtue of the transformation that would occur. Crucially for Paul, this means that standard decision procedures can't be used in cases of transformation since subjective values can't be assigned to outcomes involving transformation.

Paul's story is controversial (see Pettigrew, "Transformative Experience and Decision Theory"; Krishnamurthy, "We Can Make Rational Decisions to Have a Child"; and Dougherty, Horowitz, and Sliwa, "Expecting the Unexpected"), but this paper can sidestep the issues subjective values present. Here's why. Suppose that Paul is right that (i) agents cannot assign subjective values to futures involving personal transformation, and (ii) subjective values are what matter for standard decision theory. It then follows that standard decision theory—and thus the Pascalian model!—doesn't apply to decisions involving transformation. (If the reader is inclined towards this view, they can simply skip to §5.) But suppose that Paul is not right about subjective values and there are values relevant for self-interest. It may very well turn out (as I'll argue) that even when these values are granted, the Pascalian model still doesn't work when it comes to adopting the life of faith and decisions involving transformation.

[21]Perhaps there are exceptional cases where it is in a person's self-interest to cease existing if the only alternative to non-existence is a life that is not worth living. But this is not the situation any of our agents find themselves in, so for the purposes of this paper, we can set aside cases in which it would be better for the agent to not exist.

early life had gone better even though it would result in an agent who is radically different from who he is.

Had the past gone differently, the resulting life would be better than the one that actually has unfolded. Non-traumatized Jon experiences a great deal of value—and it's the type of value that's relevant for non-traumatized Jon's self-interest. Furthermore, unlike Wilma's child and potential child, non-traumatized Jon is metaphysically related to Jon in the right way. If self-interested value and metaphysical identity are all that are considered, then it appears that Jon is rationally required to prefer, for self-interested reasons, the alternative past over his actual past.

Nevertheless, it does not seem that Jon is rationally required to prefer that the past had gone differently. It is not hard to imagine Jon reasoning as follows: *it is true that had my early life gone better, the resulting life would be better overall than my own; however, that life would be radically different from mine, and I don't want a life so alien from my own*. There is something persuasive about this kind of reasoning, and if it's rational, then it appears that something has gone wrong with the Pascalian model. On the Pascalian model, what is more valuable for an agent is in that agent's self-interest and is the rational outcome for that agent to prefer. If that's right, then an agent cannot simultaneously hold that one outcome is more valuable for them than another and rationally prefer the worse outcome. Here, we have a case where the past sans trauma is more valuable than the past with trauma for Jon overall. Nevertheless, Jon seemingly can rationally prefer the past with trauma. Thus, we have a case that appears to cause trouble for the Pascalian model.

§3 Diagnosis

In this section, I attempt to diagnose the mismatch between value and rational preference in Jon's case, and explore the implications of that diagnosis for the Pascalian model. If we start with our observations from the case of Wilma's child, it looks like there are two ways to diagnose Jon's case. First, one might acknowledge that Jon's reasoning is persuasive, but maintain that Jon nevertheless has an irrational preference and provide an error theory for why his reasoning appears so compelling. Second, one might claim that traumatized Jon is not (metaphysically) identical to non-traumatized Jon. Then, like Wilma's child, it would be rational for non-traumatized Jon to prefer the outcome in which he is not traumatized—though things may be going extremely well for non-traumatized Jon, traumatized Jon would not exist! I'll explain why both of these diagnoses are problematic before offering my own diagnosis that makes use of self-transformation and practical identity.

3.1 Diagnosis 1: Jon is Irrational

The first diagnosis, that Jon is irrational despite the intuitive appeal of his reasoning, is unsatisfactory. It's just deeply implausible that we are

required to prefer lives that are completely foreign to our own merely because the people who have those lives experience more value than we do and happen to be metaphysically related to us. When we consider alternative ways our lives could go, we want the people in those alternatives to be people with whom we can identify. If their lives are too different from our own, they don't seem to be related to us in the right way even if the metaphysical situation is clear.

This position is not idiosyncratic. In fact, many philosophers take as a starting point the fact that from the standpoint of self-interest, agents can prefer their own lives to alternatives that are more valuable but foreign. For instance, in the context of considering the problem of evil, Adams[22] considers whether it is one's self-interest to have lived an alternative life that is absent of evil if that life would be very different from one's actual life. His paradigm case is the life of Helen Keller:

> Let us suppose that she would have had an even better and happier life if her sight and hearing had been spared (though that is not obviously true). But whatever its excellences, that life would not have had one day in it that would have been very like any day of her actual life after the age of nineteen months. Her actual life—in its emotional as well as its sensory qualities, in its skills and projects, and doubtless in much of her personality and character—was built around the fact of her blindness and deafness. That other, happier life would have contained few of the particular joys and sorrows, trials and triumphs—in short very little of the concrete content—that she cared about in her actual life. Her never having been blind or deaf would have been very like her never having existed. Why should she wish for that, given that she had reason to be glad she existed?[23]

For Adams, Keller's case shows that the link between value and rational preference can be severed. Thus, it can still be rational for Keller to prefer her own life to the alternative one since rational preference follows self-interest rather than value. Furthermore, Keller's case is analogous to Jon's. In both cases, it is rational for agents, from the standpoint of self-interest, to prefer their own lives to radically different alternatives even if those alternatives are better.

Other philosophers share Adams's view that agents can rationally prefer their own past lives to extremely different but more valuable ones—though they of course differ on *why* this is the case.[24] For instance, Harman offers an account on which we can prefer our own past to alternative ones even if it is less valuable because of our preferences for loved

---

[22]Adams, "Existence, Self-Interest, and the Problem of Evil."

[23]Adams, "Existence, Self-Interest, and the Problem of Evil," 73.

[24]For instance, the explanation for preferring the actual past Adams offers involves suggesting that there may only be partial identity between actual Helen Keller and the Helen Keller in the alternative that does not involve blindness and deafness. Harman, on the other hand, defends the past preference by arguing that there is an asymmetry between past and future desires. Those who hold "conservatism" will concur with Adams and Harman with respect to their judgment about Jon's case, but say that Jon's preference is rational because rationality requires a preference for the actual.

ones.[25] (Extending this thought to self-love also explains why Jon's pref-
erence appears rational; as long as Jon loves the person he has become,
he's reasonable in preferring his own life to that of non-traumatized Jon's.)
For our purposes, it's significant that the point of agreement is that agents
like Jon are rational when they prefer that the past went the way it did
over an alternative, but radically different possibility. And this agreement,
I suggest, counts heavily against this diagnosis.

3.2 Diagnosis 2: Metaphysical Non-identity

On the second diagnosis, Jon is not metaphysically identical to non-trauma-
tized Jon. There are at least two reasons to resist this diagnosis. First, there
is little independent motivation for claiming that claim that traumatized
Jon is not, metaphysically speaking, the same person as non-traumatized
Jon. In addition to lacking independent motivation, committing to the
non-identity of persons in divergent futures introduces a host of addi-
tional undesirable metaphysical commitments. For instance, one would
have to deny that there are other ways our lives could have gone and
hold that seemingly contingent properties are necessary.[26] Without good
independent motivation, this diagnosis is implausible.

But second, even if this diagnosis is correct, it overgeneralizes too much.
Intuitively, we could have been different. And when we imagine the ways
in which we could have been different, the difference arises from situa-
tions like the past going a different way. If agents in divergent potential
outcomes are not identical, then when it comes to an agent looking back at
ways the past might have gone differently, it is never the case that things
could have gone better for that agent. Over-caffeinated agents should be
able to look back at their past decision and prefer that they had not con-
sumed one cup too many. But they cannot do so if the alternative outcome
in which an extremely similar agent who shares their history does not
actually contain an agent who is metaphysically identical to them. Note
that this second diagnosis does not merely amount to a preference for the
actual, which is in itself a dubious result for practical rationality (and one

---

[25]While discussing the merits of "I'll be glad I did it" reasoning, Harman considers cases
in which adults are glad that their lives turned out the way that they did even though alter-
native lives would have been more valuable to the person living those lives. For instance,
a teenage mother can rationally prefer not having waited to have a child later in life and a
deaf adult can rationally prefer not having received a cochlear implant as a child even if it is
stipulated in both cases that the alternative pasts would have been more valuable. Harman
explains that this is because "[i]t can be reasonable to prefer that someone one loves has
come to be the person she is, even if one recognizes that there is an alternative in which
things would have been better, in every way of being better which one should care about. It
can be reasonable to prefer that someone one loves has come to exist" (Harman, "'I'll Be Glad
I Did It' Reasoning," 186).

[26]It would be odd if pre-traumatized Jon stood in the personal identity relation to both
traumatized Jon and non-traumatized Jon, but that traumatized Jon and non-traumatized
Jon did not stand in that relation to each other. I'm not saying that it's impossible to tell a
story that explains this oddness, but that story does have the consequence that traumatized
Jon's life could not have gone otherwise! Properties, such as being traumatized, that we
might have thought were contingent are actually necessary.

that is incompatible with preferring that your over-caffeinated self had not consumed that last cup of coffee). It requires accepting an extremely unsavory metaphysical consequence—we can only exist in the exact outcomes that played out. There is no sense in which things could have gone differently for us.

I suspect that this diagnosis is antecedently unattractive to most philosophers. But let me briefly highlight something in case the reader does find this diagnosis appealing. Should it turn out that Jon and non-traumatized Jon are not metaphysically identical, then Jon's case is analogous to that of Wilma's child. It then turns out that it is in Jon's self-interest to have experienced the childhood trauma and that he is rational in preferring that past to the non-traumatic alternative. As we'll see when we turn to Blaise and cases involving future change, this diagnosis also spells trouble for the Pascalian model.

3.3 Diagnosis 3: Practical Non-identity

Let's take stock of the ground we've covered. In rejecting the first two diagnoses, the following two commitments were affirmed:

(i) It is rational, from the standpoint of self-interest, for Jon to prefer his own life to that of non-traumatized Jon even if non-traumatized Jon experiences more value (of the kind relevant to self-interest) than Jon does.

(ii) Jon is identical (metaphysically speaking) to non-traumatized Jon.

Accepting these two commitments means that knowing all the facts about self-interested value as well as all the metaphysical facts in some outcomes is not always enough to determine which of those two outcomes the agent ought to rationally prefer. This is enough to cause a problem for the Pascalian model since considerations of self-interest are supposed to be sufficient for determining rational preference. But we've now seen that in cases like Jon's involving past decision points where an alternative outcome involves a radically different life, the links between self-interested value and rational preference appear not to hold. This means that agents like Jon, who are fully aware of all of the relevant value facts, cannot move from self-interested value to rational preference in the way the Pascalian model suggests that they can.

Thus far, I've argued against two diagnoses that would preserve the Pascalian model as typically understood. I now want to offer my own diagnosis, which can be taken as either a rejection of or modification to the Pascalian model. Notice that Jon's reasoning is persuasive because *Past trauma* is similar to *Non-identity*. In both cases, an agent compares two scenarios, one that involves who they are, and one that involves an alternative scenario with a radically different agent. In *Non-identity*, the agents in the two outcomes are not metaphysically related to each other. In *Past trauma*, the two agents are metaphysically related, but there's been significant divergence involving loss of features with which Jon strongly

identifies. The way in which he overcame his past trauma has been incorporated into how he identifies as an agent, and constitutes who he is. When he considers the alternative non-traumatized life, he can't identify with that alternative agent because that agent's life is so different from his own. As far as Jon's practical reasoning is concerned, that agent may as well be someone else—it is *as if* they are not metaphysically identical. Non-traumatized Jon is not *practically identical* to Jon.

Non-traumatized Jon lacks features that Jon takes to be practically essential to himself, so Jon cannot identify with non-traumatized Jon. In Paul's language, Jon would have to undergo a transformative experience to become non-traumatized Jon. Jon cannot project what it would be like to be non-traumatized Jon or affectively empathize with non-traumatized Jon in the way that is required to form preferences about the alternative in which he does not experience childhood trauma. In Velleman's language, Jon cannot imagine what it would be like for him to be non-traumatized Jon in the same way that we cannot imagine what it would be like for us to be Napoleon looking out at Austerlitz. Because the alternative is not one involving an agent with whom Jon identifies, the value experienced by the agent in that alternative is irrelevant to Jon when Jon considers things from the standpoint of self-interest.

Here's my diagnosis. In mundane cases, knowing all the facts about how valuable outcomes are for agents and whether those agents metaphysically exist in those situations are sufficient for determining which outcome an agent ought to rationally prefer from the standpoint of self-interest. This is because the outcomes involving the same metaphysical agent and outcomes involving the same practical agent are coextensive in mundane cases. But in cases involving self-transformation, such as Jon's, those two types of facts aren't sufficient for determining which outcome the agent ought to prefer because metaphysical and practical identity come apart. Agents must also identify in a practical sense with the agents in the outcomes. The 'self' of self-interest is often taken for granted when people use standard decision procedures, such as the Pascalian model. To the extent that it's considered, such as in non-identity cases, metaphysical identity is the default. What I'm suggesting is that if we don't want to reject the Pascalian model outright in light of cases like Jon's, it needs to be modified such that the self in 'self-interest' is the practical self. Then, Jon's preference for his actual past is a preference for the past involving the most self-interested value since the value associated with the alternative past is no longer the type of value relevant to (practical) self-interest. Optimistically for the Pascalian model, the value, metaphysical, and practical facts about the agent in all the outcomes are sufficient for determining which outcome agents are rationally required to prefer from the standpoint of self-interest.

Here's where this diagnosis leaves us. In cases involving self-transformation in the past, rational preferences depend upon self-interested value and practical (and metaphysical) identity. In §4, I'll argue that the same

holds for decisions involving future self-transformation. In §5, I'll discuss the implications of this view for the practical rationality of faith.

§4 Future-oriented Preferences

Thus far, I've argued that in cases involving past self-transformation, the Pascalian model needs to understand the self of self-interest in terms of practical identity. The same, I now argue, holds for preferences about future outcomes involving self-transformation. The argument is simple:

(P1) An agent's rational preference about past alternatives depends on (i) the value for agents in those alternatives, and (ii) the practical (and metaphysical) identity of agents in those alternatives being that of the agent deliberating about their preferences regarding the past.[27]

(P2) What the rationality of a preference depends on is insensitive to whether the preference is about the agent's past, present, or future.

(C) An agent's rational preference about future alternatives depends on (i) the value for agents in those alternatives, and (ii) the practical (and metaphysical) identity of agents in those alternatives being that of the agent deliberating about their preferences regarding the future.

I defended (P1) in the previous section. This section is dedicated to defending (P2).

Prima facie, it seems that whatever the rationality of past preferences depends upon is what present and future preferences also depend upon. (P2) is a reasonable default position. It's even plausible to interpret the Pascalian model as holding this position as well—an agent's rational preferences are linked to self-interested value for that agent, regardless of whether that preference is about the past, present, or the future. Furthermore, there are a variety of arguments defending treating past, present, and future preferences similarly. For instance, the Epicureans thought that differences based on whether the preference is about the past, present, or future, are arbitrary ones that rationality ignores.[28] Sullivan and Greene

---

[27]Alternatively, if the reader prefers to reject the Pascalian model rather than adopt the modification involving practical identity, the reader can use something like (P1*):

(P1*) It's not the case that the Pascalian model determines a self-interested agent's rational preference about past alternatives in cases of self-transformation.

(P2) remains the same, and combined with (P1*) yields the alternative (C*):

(C*) It's not the case that the Pascalian model determines a self-interested agent's rational preference about future alternatives in cases of self-transformation.

[28]On the Epicurean view, self-interest is a matter of pleasure—the self-interested agent ought to maximize the amount of pleasure in his or her life. We can think of the amount of pleasure an agent experiences during any given event as the subjective value that event has for the agent. What the Epicureans realized was that tradeoffs in pleasure often have to be made between near- and distant-future events, and that one's life ends up having more total value when near- and distant-future events are weighted equally.

defend this line by pointing out that when agents treat past preferences differently by discounting them, agents can be Dutch-booked, which demonstrates that their preferences are inconsistent, and thus irrational.[29]

However, (P2) is not completely unassailable. There are some asymmetries between the past and the future. Perhaps one of these asymmetries affects (P2). I'll briefly consider two such asymmetries and the objections to which they give rise. First, one might hold that the past is importantly different from the future because while the past is fixed (at least with respect to its intrinsic features), the future is open.[30] Because of this asymmetry, agents can't choose an alternative past while they can choose amongst alternative futures. More problematically—or so the objection goes—agents can't form preferences about alternatives (such as past ones) that they cannot actually choose. But as we saw earlier, the claim that we cannot form preferences and make choices with respect to the past in light of its fixedness is dubious. Though the asymmetry with respect to fixedness might affect which options are live—Jon can't do anything to change his past but has options with respect to the future—it does not affect the ability to form rational preferences. Thus, even if there is this asymmetry between the past and the future, it does not affect an agent's ability to form preferences about the past.

Second, one might object to (P2) because of a causal asymmetry between past and future states of affairs. Past states cause future states, but not vice versa. Part of why one might be inclined to believe that Jon can prefer his actual past is that the actual past (or something similar that would include the trauma) is causally necessary for his becoming who he is. Jon, who is happy with who he has become, cannot prefer a significantly different past because that would entail radically changing who he has become. However, what happens in the future has no causal effect on who he is now. Thus, the objection goes, one can prefer a past with less value than an alternative if it leads to the present that is preferred.[31] But preferring a future outcome with less value than an alternative cannot be defended on the same grounds since the future outcome does not causally affect the present.[32]

---

[29]Sullivan and Greene are responding to Parfit. Parfit presents thought experiments that purportedly show that it is rational for agents to prefer that unpleasant events be in their past rather than their future. Sullivan and Greene point out that assigning any kind of discount function to the past, one becomes vulnerable to Dutch-booking. Because Dutch books occur when there are inconsistent value and probability assignments, avoiding Dutch books is often taken to be a necessary condition for rationality. See Greene and Sullivan, "Against Time Bias."

[30]The metaphysics underlying this point is controversial, but this paper grants it for the sake of argument.

[31]One can also think of the preference for the actual and its associated value as a type of conservatism about value. For more discussion on (the implausibility) of this view, see Finocchiaro and Sullivan, "Yet Another 'Epicurean' Argument," especially section 5.5.

[32]E.g., Harman, "'I'll Be Glad I Did it' Reasoning." On her view, the young parent and deaf adult can prefer their actual pasts because of the personal attachments they value. However, the teenager who is a merely potential young parent and the deaf child should not have the

In addressing this objection, it's important to return to why Jon prefers his actual past involving the trauma to the one in which he does not experience the trauma. Jon doesn't prefer his actual past merely because it is the only past that could result in who he is; he prefers it because he regards it as an integral part of who he is. His connection with that particular past is more than merely causal. His past *constitutes* who he is now, and has become a practically essential feature of who he is; it's integrated into his practical identity. The causal and constitutive aren't easily pulled apart, but carefully thinking about pairs of cases can help us better understand the distinction. Imagine twin-Jon, who is like Jon in every respect except for the way in which he identifies with his past. Jon's past has informed the person he has become, and is integrated into the person into whom he has developed. While twin-Jon has followed the same path as Jon, he resents his traumatic past and wishes that it had not happened. Like Jon, twin-Jon likes the person he has become, would not necessarily prefer to be someone radically different, and realizes that his past has causally led to who he is. However, unlike Jon, twin-Jon resents his past. Twin-Jon does not regard his past as integrated into who he is. If there were a miracle world in which he developed into a similar person without the traumatic past—a world in which the causal connection is severed—twin-Jon would prefer that world to the actual one.[33] But Jon would not because his past is a practically essential part of who he is. While twin-Jon prefers his own past to competing alternatives in virtue of only the causal connection, Jon prefers his past in virtue of it being constitutive of who he is.

This distinction reveals why the causal asymmetry objection appears convincing. If we have in mind preferences that are formed in virtue of the merely causal connection, then (P2) appears false since the causal connection does not apply to future states. However, if the preference is based in the past's being constitutive or essential to who the agent is, then this objection doesn't demonstrate the falsity of (P2). Instead, the objection would need to rely upon an asymmetry between the past and future when it comes to being a constitutive or essential part of who a person becomes. But once we shift to constitution rather than causation, it is not clear that there is an asymmetry between the past and the future. Instead, if a potential future lacks a feature that is practically essential to the agent, then it seems that the deliberating agent is rational in avoiding that future even if it is more valuable than one for which the agent ultimately opts. The agent does not exist in that future, practically speaking. So unlike causation, constitution does not appear to give rise to an asymmetry be-

---

corresponding future preferences that involve young parenthood and a lifetime of deafness because they have not yet formed the personal attachments that justify the backward-looking past preference.

[33]For instance, on the Hudson view of hypertime, one can imagine Jon and twin-Jon comparing the actual world and a descendant of the actual world where God has removed the portion of the four-dimensional spacetime block containing the trauma (Hudson, *The Fall and Hypertime*).

cause it's connected to the essential features of an agent's practical identity. Indeed, considerations regarding these essential features of practical identity seem to give a very good reason for thinking that future preferences work similarly to past ones. Thus, the objection to (P2) fails.

If what I've argued for here is right, then rational preferences about the future take practical identity (as well as metaphysical identity and value) into account. We can now return to Blaise, who is considering whether to take up the deity on the offer to become the happiest man in the world. Taking the offer requires self-transformation, so Blaise is not identical, practically speaking, to the unsurpassably happy agent. Thus, Blaise is not rationally required to prefer that outcome even though it has an agent who experiences a great deal of value and is metaphysically related to Blaise. The same holds for the potential parent, young Russian, and any other agent facing future self-transformation. As with past preferences, if the Pascalian model is to work, it must take into account practical identity. But once it does, it looks like from the standpoint of self-interest, agents aren't rationally required to prefer futures in which they undergo self-transformation.

§5 Faith as Transformative

5.1 The Transformative Nature of Faith

Agents who are considering adopting the life of the religiously faithful also face an outcome in which they undergo self-transformation. Faith is transformative, and there are many motivations for thinking that it fits the paradigm of self-transformation. The Apostle Paul, who undergoes a rather striking conversion, is in a very real sense a different person from Saul. To be sure, Saul doesn't *choose* to convert; rather, conversion is thrust upon him via an unexpected religious experience. However, his conversion does demonstrate the transformation that occurs as a result of conversion. Saul's core preferences and life goals radically change. Had they been cotemporaneous, his pre-transformation self would have persecuted his transformed self. The way he sees and experiences the world is wildly different as well: it is as if scales have fallen from his eyes. Though Saul did not choose to be transformed, we can easily imagine his earlier self being horrified at the thought of choosing that future. From the perspective of Paul pre-transformation, the value associated with the apostolic life is does not affect his self-interest since the apostolic future self may as well be a completely distinct person.

Paul's case is extreme, but there are strong reasons for thinking that adopting the life of faith requires a similar degree of radical personal change. First, part of adopting this life traditionally involves submitting oneself to God's will. This submission involves a deep change in core preferences and life goals. For instance, from the Christian tradition, baptism symbolizes rebirth into a new life of faith. In the New Testament, Jesus tells a potential follower to "let the dead bury the dead" (Matthew 8:22).

Followers must be willing to radically change core aspects of their lives and concerns. Second, part of adopting the life of faith is hope for the afterlife. But the afterlife is almost certainly transformative. An eternity spent enjoying the beatific vision is like nothing else experienced in this world. As Sullivan[34] points out, in light of how foreign that future would be, it is not obvious that one ought to prefer to go on. Adopting the life of faith is well within the paradigm of self-transformation.

As these examples suggest, faith—or at least the species of faith that I focus on—involves action.[35] Namely, it involves adopting and then living the life of faith. One makes a lifelong commitment to continually *act* in a particular way. For instance, if a core commandment involves loving others as oneself, living faithfully requires continually striving to love others. (For the purposes of the paper, it doesn't matter what specifically constitutes the life, so long as it is sufficiently different from the life the agent would live without faith.) The lifelong nature of religious commitment is not unlike the commitment to a spouse; both demand living substantially different than one would otherwise, and carry on indefinitely with the intent that they be irreversible. Carrying out this commitment is essentially action-based, so choosing to be transformed by faith falls squarely into the domain of practical rationality. Whether agents are rationally required to prefer and choose to adopt this life depends upon the considerations discussed above with respect to self-transformation.

Finally, it's worth noting that it is compatible with what I've said that living the life of faith is a way of acquiring propositional faith (i.e., "faith that"). However, in keeping with Pascal and others, I'll take for granted that one can live the life of faith independently of whether one has propositional faith.[36] Recall that after giving the wager, Pascal considers whether agents have enough voluntary control over their beliefs to choose to adopt belief in God. He then suggests that agents can adopt a way of life that may eventually lead to belief.[37] Regardless of whether Pascal's suggestion

---

[34]Though Sullivan ultimately thinks that one can rationally prefer to go on, she also argues that it's not *obvious* that one should given how transformative the afterlife might be. See Sullivan, *Time Bias*, chap. 10 "Neutrality and Life Extension."

[35]It's widely accepted that there are multiple species of faith. Audi, for instance, proposes that there are at least seven distinct species of faith ("Faith, Belief, and Rationality"). The species I'm concerned with involves action. For contrast, there are also species that are propositional, or take the form of "faith that p" where p is some proposition. Those species are outside the scope of this paper, though I leave open the possibility that they might be *expressed* by actions. If they are, then perhaps what I say here also bears on other species of faith.

[36]I admit that this is controversial since one might think that acting faithfully requires having epistemic faith in certain propositions. However, it seems entirely possible that one can mimic what it is to live the religious life, and that doing so will also require commitment that rises to the level of being transformative.

[37]Recall that in the *Penseés*, Pascal considers the person who might object as follows: "I am made so that I cannot believe. What then do you wish me to do?" Pascal advises that those who find themselves in this position should try "acting as if they already believed, taking holy water, having masses said, etc."

is an efficacious way of gaining propositional belief, it is clear that one can choose to commit to the life of faith independently of one's propositional commitments. But should it turn out that acting as if one believes does lead to belief, that is yet another way in which faith is potentially transformative.

Given these features of living the life of faith, one might wonder if faith is universally transformative. For instance, one might think that what I've said here doesn't apply to a person who has already adopted the life of faith but is deliberating about whether to abandon it. If the person has already adopted that life, then living faithfully already figures into their practical identity, and continuing to remain committed to it wouldn't involve self-transformation. Indeed, abandoning it might be the option that involves self-transformation if the agent's faith has been essential to their self-identity. Though this is clearly not the type of agent Pascal had in mind, the case poses an interesting question about the universality of self-transformation in the case of faith. Similarly, an agent who is sufficiently "easy-going" and doesn't self-identify with any features might not risk losing any practically essential features by adopting a life of faith.[38] (If this is right, then it turns out that the Pascalian model might rationally require an agent to live the life of faith in this type of situation.)

I agree that based on what's been said about self-transformation, it is possible that there are agents who would not regard adopting (or continuing) the life of faith as transformative. I leave it to the reader to decide for themselves how many people are in this situation. However, there are two things worth noting. First, what these cases demonstrate is that a person's practically essential features factor into their evaluation of the potential outcomes. This reinforces my central point that the Pascalian model must take practical identity into account. Second, there very well may be aspects of the life of faith, such as the beatific vision, that are universally transformative. Given that it involves changes we can't comprehend and that we have at least some practically essential features, it seems possible that the changes might involve loss of those features. I'm not prepared to fully argue for this point, but it seems plausible. And even if there are not aspects of faith that are universally transformative, it does seem that for most people, adopting the life of faith would involve self-transformation.

5.2 Consequences for Faith

The transformative nature of faith, combined with the discussion in the previous sections, suggests that the Pascalian model of reasoning does not require deliberating agents to adopt a life of faith. This result has some interesting consequences. First, Pascal's wager fails. On a Pascalian model that assumes that only metaphysical identity and value matter for self-interest, we run into the future version of *Past trauma*. Then just as Jon is not rationally required to prefer the alternative past involving an agent

---

[38]Thanks to two anonymous referees for this point.

with which he cannot identify, the potential faithful agent is not rationally required to prefer a future outcome involving a transformed agent with which the deliberating agent does not identify. On the revised Pascalian model that uses practical identity as what matters for self-interest, it turns out that the future agent who has been transformed by faith is not practically identical to the deliberating agent. Thus, the deliberating agent is not rationally required to prefer the future outcome involving the life of faith.

This consequence is significant, especially in light of how much of the literature surrounding the wager revolves around issues involving probability and infinite value.[39] What my argument shows is that whether the wager can be formulated to avoid these problems is not what matters. Even if the math works out in favor of Pascal, we've seen that the fact that there is an outcome involving maximal value for the agent in that outcome does not entail that a deliberating agent is rationally required to prefer that outcome from the standpoint of self-interest. For if that outcome involves self-transformation, the deliberating agent may not be practically related to the transformed agent, and the self-interest relation won't hold. Agents aren't rationally required to take the wager since the infinite payout belongs to an agent that is not practically identical to them.

Second, my argument shows that the related "Anselm's wager" does not rationally require agents to adopt the life of faith either. On Anselm's wager, the deliberating agent knows that God exists and that the life of faith holds infinite value, so the element of risk present in Pascal's wager is eliminated. The transparency of Anselm's wager is even closer to Blaise's situation than Pascal's wager is; the agent simply chooses whether they want to adopt the life of faith. Nevertheless, from the standpoint of self-interest, an agent facing this choice still isn't rationally required to prefer the outcome with infinite value since, practically speaking, they are not the agent enjoying that value.

Third, my argument bears on contemporary attempts to justify the rationality of religious faith by appealing to value. Like Pascal's wager, this work on the rationality of faith also relies on faith being practically rational. For instance, Buchak offers an analysis of faith on which "Act A is an act of faith that X, for S, if and only if: (i) S performing act A constitutes S taking a (subjective) risk on X, (ii) S chooses (to perform A before he examines additional evidence) rather than (to postpone his decision about A until he examines additional evidence), and (iii) S would follow through on performing A even if he were to receive counter-evidence."[40] Initially, it might appear that both (ii) and (iii) potentially require one to act contrary to what is practically rational. First with respect to (ii), it's generally

---

[39]Hájek, "Waging War on Pascal's Wager," and Monton, "Mixed Strategies Can't Evade Pascal's Wager," for instance, argue about whether mixed strategies show that the wager does not work. Even more recently, Michael Rota ("A Better Version of Pascal's Wager") has argued that a better version of the wager can be formulated to avoid problems related to infinite utilities and the many-gods objection.

[40]Buchak, "Faith and Steadfastness," 120.

practically better for an agent to have more rather than less evidence when making a decision. Second, with respect to (iii), it seems that counter-evidence can sometimes lower one's credence in X enough to require one to refrain from acting on X.

However, it turns out that under certain conditions, it can actually be rational for an agent to have faith. First, it turns out that there are two scenarios in which the generalization about additional evidence does not apply—scenarios in which additional evidence is too costly to obtain and scenarios in which the agent will perform the same action regardless of the evidence.[41] Plausibly, both of these scenarios are in play in the context of faith. Buchak suggests that postponing the decision and examining the evidence may involve too high a cost in some interpersonal contexts, including religious ones, since it risks damaging a relationship where faith in the other person matters. Also, condition (iii) suggests that the faithful agent would perform act A even in the face of counter-evidence. Second, commitment in the face of counter-evidence can be required to obtain a long-term good—without it, we might too hastily abandon projects when presented with misleading evidence. The value of the long-term good might then be weighty enough to make the risk of performing act A even in the face of counter-evidence practically rational. This is especially important in the contest of faith since faith takes place over a long time-scale during which one might receive additional evidence even if one is attempting to not examine further.

While Buchak's formulation is a general account of faith, it's also applicable to religious faith. All that's needed to apply it to religious faith is the relevant religious content and then an act that expresses that religious commitment, such as adopting the life of faith. Thus, if Buchak is correct, then she's demonstrated that faith—both generally and in the religious case—can be justified by appealing to practical rationality. Now there's a further question of whether the values in play in Buchak's account are self-interested values. If they are, then her account is one that follows the Pascalian model, and, in light of what I've said above, agents are not rationally required to perform acts that express faith. However, if some of the values are not—perhaps because faith is objectively intrinsically valuable or has a moral dimension—then agents might still be rationally required to perform acts of faith. In any case, it's significant that the model does not work if self-interested values are the only ones in play.

McKaughan[42] also offers an account on which faith is practically rational. He distinguishes between different accounts of faith, and points out that on at least two models—models on which faith is either trust-based or hope-based—faith does not require belief in some proposition. If this is the case, then faith needs to be evaluated in accordance with non-epis-

---

[41]For more detailed discussion of when it is rational to forego additional evidence, see Buchak, "Instrumental Rationality, Epistemic Rationality, and Evidence-Gathering."

[42]McKaughan, "Authentic Faith and Acknowledged Risk."

temic, rather than epistemic, considerations. These considerations turn out to be practical ones. McKaughan's suggestion is that though having trust- or hope-based faith involves risk, it can be rational to take on that risk if the potential benefits are high enough. The benefits or value of the outcomes for the agent is what grounds the rationality of faith. Thus, what I say above about Buchak's overall account applies here as well.

In a later paper, McKaughan[43] offers an action-centered account of faith that appeals to standard decision theory. Interestingly, the values in play in this account arise, in some sense, from the agent. Here, the agent endows the outcome involving faith with more value by subjectively valuing it. Valuing this outcome is not itself required by practical rationality—so unlike Pascal, choosing the life of faith ends up being permissive rather than rationally required. However, once the agent places value on that outcome, the outcome becomes valuable, which then makes the agent's preference for that outcome practically rational. I find this account interesting because since it's a permissive account, it's compatible with my claim that self-interest alone can't rationally require an agent to adopt the life of faith. But because faith is transformative, there's still a question of whether the self in the outcome involving faith is practically identical to the self who places value on that outcome. If not, then it will still turn out that the expected value (if it is self-interested value) still may not be sufficient to make the agent's choice rational even if the agent places value on the outcome involving faith.

Buchak and McKaughan are not alone in attempting to defend the rationality of faith by appealing to practical benefits. Audi[44] also argues that some forms of faith do not require belief and that whether those forms of faith are rational does not turn on whether certain theistic beliefs are true or justified. The central theme in all of these defenses of faith has a Pascalian ring to it. Like Pascal's wager, these contemporary defenses of faith appeal to the value an agent stands to gain if he or she adopts the life of faith. The working assumption is that the game is over once it's established that the outcome in which the agent has faith has more value for the agent than the outcome in which the agent lacks faith. However, if my argument above is correct, these defenses are not sufficient for showing that it is in an agent's self-interest to have religious faith since it involves self-transformation. Though it is true that there is more value for the agent in the outcome that involves faith, the agent in that outcome may not be related to the deliberating agent in the correct way since faith is transformative. Thus, from the standpoint of self-interest, the deliberating agent is not rationally required to prefer adopting the life of faith.

---

[43]McKaughan, "Action-Centered Faith, Doubt, and Rationality."

[44]Audi, "Faith, Belief, and Rationality."

§6 Concluding Thoughts

I have argued that the Pascalian model of reasoning doesn't work as expected in cases involving self-transformation. I then suggested that adopting a life of faith is a case of self-transformation, so, *a fortiori*, the Pascalian model doesn't end up rationally requiring agents to adopt a life of faith even if that life involves infinite value for an agent. As a result, Pascal's wager, and any other attempt to defend the practical rationality of faith solely in terms of self-interest is bound to fail. In the remainder of this paper, I want to consider three upshots of my view.

First, the points made above provide a new way of understanding the "wrong reasons" objection to Pascal's wager. On the canonical understanding of this objection, taking the wager on the basis of self-interest is a wrong reason—it comes from an ignoble motivation. It's been noted that this motivation is problematic for at least two reasons. First, it's dubious that God would reward those who have wrong reasons with infinite value. If it turns out that God doesn't reward those with wrong reasons, then that motivation for taking the wager is undercut. Second, one might be troubled by having the prospect of infinite value for oneself being the motivation for living the life of faith. Presumably, one of the most important parts of faith is seeking a personal relationship with God, and basing one's faith solely on value misses this point. Doing so would be as disingenuous as marrying someone solely because they are rich. Ideally, one ought to be motivated to seek a relationship with another person because they care about that person, not because they care purely about their self-interest. Understood in this way, the wrong reasons objection highlights a sort of interpersonal moral failing that arises when value for oneself, rather than genuine care for God, grounds one's faith.

On my account, making the decision because of the infinite payout associated with the outcome is not merely potentially undercutting or an interpersonal moral failing—it is also troubling from the standpoint of self-interest. After all, the agent enjoying that infinite value may not be related to the deliberating agent in the way relevant for self-interest! It turns out that value is a "wrong reason" in an additional sense: it is irrational for an agent to prefer an outcome on the basis of value if the agent also fails to consider whether they exist, practically speaking, in that outcome.

Second, my account has an interesting consequence for one response to the problem of divine hiddenness. One justification for hiddenness involves the need for distance between human agents and God so that human agents can freely decide whether they want to seek a relationship with God. On this response to hiddenness, God is irresistible. God is so irresistible that were God to be fully revealed, humans would be so overwhelmed that they would not be able to turn away from a relationship. Because they would not be able to do otherwise, the preconditions for freely choosing would fail to be met.

If I'm right, then it seems possible that human agents could resist and do so rationally. God could be fully revealed, and we would find ourselves in an Anselm's wager situation. Nevertheless, because committing to the life of faith would involve self-transformation, the value experienced by the agent who transforms by committing to that life would not be a self-interested reason for the deliberating agent to commit. We would find ourselves in Blaise's situation where it would be possible to reject the life of faith, and reject it rationally. Thus, this justification for God's hiddenness fails.

Finally, given what I've said, one might wonder whether agents can rationally prefer to adopt the life of faith given that it involves self-transformation. I've argued that in the scenario Pascal envisions, agents are not rationally required to adopt the life of faith. I've left open whether agents can rationally prefer that life at all. Perhaps they are rationally required not to prefer that life, or perhaps that preference is permissive (i.e., it is rationally permissible to prefer that life and rationally permissible not to have that preference). Both of these possibilities are interesting because they make room for a practical version of fideism. Fideism is typically an epistemic position on which, roughly, agents who have faith that P (where P is some proposition) believe P and are not rationally required to believe that P. On a more extreme version of fideism, belief that P is irrational, and on a less extreme version, belief that P is permissive. Practical fideism would then be the view that agents who have faith expressed by some preference (or corresponding choice) for a state of affairs S are not rationally required to prefer S. Again, on a more extreme version, preferring S is irrational, and on a less extreme version, preferring S is permissive.

I've also left open whether there are other factors that might require agents to adopt that life. The Pascalian model takes only self-interested value into account. But there might very well be other factors in play. Some of these potential factors, such as moral ones, go against the Pascalian spirit of trying to convince agents by appealing to self-interest. But my account leaves open the possibility that such factors might exist and affect what is practically rational for agents. Thus, it's possible that practical rationality that factors in more than self-interest might require agents to adopt the life of faith. There also might be considerations other than self-interested value or moral ones that are connected to practical rationality. There isn't room in this paper to fully pursue what these non-evaluative considerations might be, but this possibility opens the door to some interesting candidates. For instance, in the literature on reasons, some have argued that voluntarist reasons (i.e., reasons arising from an act of will on the part of the agent) can provide the grounds of rational preference in some cases.[45] These will-based reasons don't reduce to self-interested value, and

---

[45]For instance, Chang ("Grounding Practical Normativity") and Korsgaard (*The Sources of Normativity*") argue that acts of the will can generate reasons for action. These reasons aren't based on self-interested value—value considerations are already taken into account as given reasons that weigh into agents' deliberation. See also Chang, "Transformative

perhaps could make room for a different type of account of the practical rationality of faith.

Unfortunately, there isn't room in this paper to fully pursue either of these possibilities. However, I've raised them to highlight where this paper leaves the discussion. These areas of future exploration are especially interesting because they also apply more broadly to other cases of self-transformation.

At the end of the passage in the *Penseés* containing his wager, Pascal imagines his interlocutor having a final reservation. His interlocutor has just learned that taking the wager is (purportedly) in his self-interest and is now reflecting on Pascal's claim that that action (i.e., adopting the life of faith) can induce the belief required by the wager. The interlocutor, mulling over taking the prescribed action and the changes that action will bring about, finally says "but that is what I fear." Pascal responds by asking "Why? What do you have to lose?" and goes on to enumerate the ways in which the interlocutor will change. Pascal concludes the passage by painting a picture of an interlocutor whose fears have been assuaged. I hope that the discussion in this paper reveals that the interlocutor's fears are indeed apt and provides an overlooked response to accounts of faith that rely upon self-interest and the Pascalian model.[46]

San Jose State University

References

Adams, Robert. 1987. "Existence, Self-Interest, and the Problem of Evil." In *The Virtue of Faith* (Oxford University Press), 65–76. https://doi.org/10.2307/2214096

Audi, Robert. 1991. "Faith, Belief, and Rationality." *Philosophical Perspectives* 5: 213–239. https://doi.org/10.2307/2214096

Barnes, Elizabeth. 2015. "Social Identities and Transformative Experience." *Res Philosophica* 92, no. 2: 171–187. https://doi.org/10.11612/resphil.2015.92.2.3

Buchak, Lara. 2010. "Instrumental Rationality, Epistemic Rationality, and Evidence-Gathering." *Philosophical Perspectives* 24: 85–120.
https://doi.org/10.1111/j.1520-8583.2010.00186.x

Choices," and Chan, "Religious Experience, Voluntarist Reasons, and the Transformative Experience Puzzle," for suggestions on how voluntarist reasons might provide a way to rationally choose to undergo transformation.

Buchak, Lara. 2012. "Can it Be Rational to Have Faith?" In *Probability in the Philosophy of Religion*, edited by Jake Chandler and Victoria S. Harrison (Oxford University Press), 225–247.
https://doi.org/10.1093/acprof:oso/9780199604760.003.0012

Buchak, Lara. 2017. "Faith and Steadfastness in the Face of Counter-evidence." *International Journal of Philosophy of Religion* 81: 113–133.
https://doi.org/10.1007/s11153-016-9609-7

Chan, Rebecca. 2016. "Religious Experience, Voluntarist Reasons, and the Transformative Experience Puzzle." *Res Philosophica* 93, no. 1: 269–287.
https://doi.org/10.11612/resphil.2016.93.1.16

Chang, Ruth. 2013. "Grounding Practical Normativity." *Philosophical Studies* 164, no. 1: 163–187. https://doi.org/10.1007/s11098-013-0092-z

Chang, Ruth. 2015. "Transformative Choices." *Res Philosophica* 92, no. 2: 237–282.
https://doi.org/10.11612/resphil.2015.92.2.14

Dougherty, Tom, Sophie Horowitz, and Paulina Sliwa. 2015. "Expecting the Unexpected." *Res Philosophica* 92, no. 2: 301–321.
https://doi.org/10.11612/resphil.2015.92.2.5

Finocchiaro, Peter, and Meghan Sullivan. 2016. "Yet Another 'Epicurean' Argument." *Philosophical Perspectives* 30, no. 1: 135–159.
https://doi.org/10.1111/phpe.12081

Greene, Preston, and Meghan Sullivan. 2015. "Against Time Bias." *Ethics* 125, no. 4: 947–970. https://doi.org/10.1086/680910

Hájek, Alan. 2003. "Waging War on Pascal's Wager." *Philosophical Review* 112, no. 1: 27–56. https://doi.org/10.1215/00318108-112-1-27

Harman, Elizabeth. 2009. "'I'll Be Glad I Did it' Reasoning and the Significance of Future Desires." *Philosophical Perspectives* 23, no. 1.
https://doi.org/10.1111/j.1520-8583.2009.00166.x

Harman, Elizabeth. 2011. "Fischer and Lamenting Nonexistence." *Social Theory and Practice* 37, no. 1: 129–142. https://doi.org/10.5840/soctheorpract20113719

Harman, Elizabeth. 2015. "Transformative Experiences and Reliance on Moral Testimony." *Res Philosophica* 92, no. 2: 323–229.
https://doi.org/10.11612/resphil.2015.92.2.8

Hudson, Hud. 2014. *The Fall and Hypertime* (Oxford University Press).
https://doi.org/10.1093/acprof:oso/9780198712695.001.0001

Korsgaard, Christine. 1996. *The Sources of Normativity* (Cambridge University Press). https://doi.org/10.1017/CBO9780511554476

Kripke, Saul. 1980. *Naming and Necessity* (Harvard University Press).

Krishnamurthy, Meena. 2015. "We Can Make Rational Decisions to Have a Child: On the Grounds for Rejecting L. A. Paul's Arguments." In *Permissible Progeny*, edited by Richard Vernon, Sarah Hannan, and Samantha Brennan (Oxford University Press). https://doi.org/10.1093/acprof:oso/9780199378111.003.0008

McKaughan, Daniel. 2013. "Authentic Faith and Acknowledged Risk: Dissolving the Problem of Faith and Reason." *Religious Studies* 49: 101–124.
https://doi.org/10.1017/S0034412512000200

33

McKaughan, Daniel. 2016. "Action-Centered Faith, Doubt, and Rationality." *Journal of Philosophical Research* 41: 71–90.
    https://doi.org/10.5840/jpr20165364

McKinnon, Rachel. 2015. "Trans*formative Experience." *Res Philosophica* 92, no. 2: 419–440. https://doi.org/10.11612/resphil.2015.92.2.12

Monton, Bradley. 2011. "Mixed Strategies Can't Evade Pascal's Wager." *Analysis* 71: 642–645. https://doi.org/10.1093/analys/anr092

Nagel, Thomas. 1986. *The View from Nowhere* (Oxford University Press).

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

Pascal, Blaise. 1670. *Penseés*. Translated by John Warrington (Dent).

Paul, L. A. 2014. *Transformative Experience* (Oxford University Press).
    https://doi.org/10.1093/acprof:oso/9780198717959.001.0001

Paul, L. A. 2015. "What You Can't Expect When You're Expecting." *Res Philosophica* 92, no. 2: 149–170. https://doi.org/10.11612/resphil.2015.92.2.1

Paul, L. A. 2016. "The Subjectively Enduring Self." In *The Routledge Handbook of the Philosophy of Temporal Experience*, edited by Ian Philips (Routledge).

Pettigrew, Richard. 2015. "Transformative Experience and Decision Theory." *Philosophy and Phenomenological Research* 91, no. 3. https://doi.org/10.1111/phpr.12240

Rota, Michael. 2016. "A Better Version of Pascal's Wager." *American Catholic Philosophical Quarterly* 90, no. 3: 415–439. https://doi.org/10.5840/acpq20166288

Schechtman, Marya. 1996. *The Constitution of Selves* (Cornell University Press).

Sullivan, Meghan. 2018. *Time Biases* (Oxford University Press).
    https://doi.org/10.1093/oso/9780198812845.001.0001

Velleman, David. 1996. "Self to Self." *The Philosophical Review* 105, no. 1: 39–76.
    https://doi.org/10.2307/2185763