An Ethics of Philosophical Belief:

The Case for Personal Conviction

Chris Ranalli

c.b.ranalli2@vu.nl

Attitude in Philosophy, Oxford University Press. Sandford Goldberg and Mark Walker (eds.)

1. Introduction

Some people believe that you shouldn't have been born (Benatar 2006). Others think that our lives are meaningless (Benatar 2017; Nagel 1971; see also Pratt 1994). Indeed, some think that you don't even exist (Unger 1979a, 1979b; Metzinger 2010), while some political thinkers of the mid 20th century think that you shouldn't exist—that you may be subjugated because of your ethnicity, sexuality, gender, class, ability, or religion, for example.[1]

These positions are 'theoretical' and 'revisionary'. What they have in common is that each one seemingly implies that many everyday beliefs are false, and that their prospective support is born out of theoretical argument, the kind of considerations one would find in an academic book or journal. A theoretical reason, a reason born out of a theoretical argument, is put forward as support for a theoretical position. And, in each case, they would have seemingly devastating consequences for many of our everyday, pre-theoretical beliefs.

What should our attitudes to revisionary theories be? Here's a plausible idea. If the arguments for these positions are evidence, then we can't simply ignore them once we're made aware of them. Whatever our attitudes ought to be, then, it's plausible that they are attitudes that should depend on our evaluation of the proffered evidence for these views. Thus, belief, disbelief, and suspension of judgment looks like the natural menu. Alternatives include varying degrees of confidence. And yet, once we factor in the theoretical evidence, we *might* no longer be justified in our initial pre-theoretical beliefs or credences. Engaging with revisionary theory risks radical change—a contrary way of thinking, or one too far away from our original confidence to recognize ourselves as holding the same values as we did before—and the risk is especially strong when it's hard for us to identify any independent fault in the theory; a fault that doesn't depend on presupposing that our pre-theoretical beliefs are justified.

Many philosophers are suspicious of the prospects of purely theoretical argument to significantly undermine our everyday beliefs—"there's chairs in the room", "there's more people

[1] For example, in the works of Evola, Gentile, and various "dark enlightenment" thinkers. Plato and Aristotle also held morally reprehensible views about how other people could be treated.

here than I expected"—and we might naturally think this extends to our deeply personal beliefs, beliefs like "I love my spouse", "My children's lives have value", "My friends are good people", and so forth. *Mooreans*, for example, will say that the premises of theoretical arguments can't be better justified than our pre-theoretical beliefs are already.[2] The revisionary philosopher's arguments about having children should have no more sway over us than the psychoanalyst's debunking arguments about love, or the social theorist's revisionary arguments about value and meaning in life.[3]

Even aside from Moorean sympathies, many in the humanities practice as if there is little, if any, potential impact from revisionary theory for everyday belief. Error theorists apparently don't suspend judgment on whether it was wrong for their loved ones to be mistreated. Nihilists apparently don't treat their friends as though they're worthless. Practically speaking, they haven't put their money where their mouth is. Survey the humanities and we see some allegiance to revisionary theories: that capitalism makes our lives effectively meaningless, that we should voluntarily go extinct, and so forth. Maybe they're inauthentic, or maybe they're just bluffing. But assuming they do believe what they purport to believe, the mismatch between their attitudes and their everyday commitments is puzzling, given that it's widely assumed that theoretical reasons can be *evidence* which bears on what we ought to believe.[4]

After all, theoreticians put forward arguments not only for the recipient to ponder, hypothesize, or make conceptual connections, but to *believe*. In a word, theory bears on belief. Indeed, this norm reflects a kind of Socratic ideal, whereby we take it that theoretical considerations ought to be considered seriously by ordinary believers once apprised of them. Socrates's street dialogues were usually with lay Athenians and not only other philosophers.

Moreover, we sometimes find people who are practically moved by theoretical considerations. Some people reorganize their lives in response to philosophical arguments about the far future, some of which are strikingly revisionary about the value of present people (Eikenberry & Mirabella 2018; Greaves & MacAskill 2021). Some argue that fundamentalism and certain strands of extremism are reactions to worries about nihilism, whether believed in oneself or others (Marty & Appleby 1994; Gilligan 2007). Undoubtedly, many religious believers sometimes come under pressure from theory as well. How could the believer, who finds meaning in their devotion, simply

[2] For developments of this approach, see Kelly (2011; 2008; 2005) and Hirsch (2002). For criticism, see Rinard (2013).

[3] Wittgensteinians or *hinge epistemologists* will say that our deeply personal beliefs have commitments that are "beyond being justified and unjustified" for purely logical reasons (Wittgenstein 1969, §359). For our purposes, this would be an overreaction to revisionary theorizing.

[4] See Pasnau (2018) for this idea generally. Many philosophers have recognized that disagreement over theoretical views can also exert epistemic pressure. See Christensen (2007).

ignore the fact that there is longstanding expert disagreement about even the existence of what one worships? While there might be ways of reconciling theory in these cases with one's practical activity, it's hard to deny the prima facie tension that theory can sometimes present.

This previews the puzzle I will explore in this paper. Theoretical arguments for revisionary theories are evidence and thereby not properly ignored. In turn, revisionary theory qua evidence simply can't be discounted. And yet we're not always in a position to know where (if at all) the theoretical evidence goes wrong. For starters, we may not know enough about the relevant domain. Indeed, due to specialization, it will be increasingly difficult to identify a flaw in a revisionary argument in another area of specialization. I had children, but I don't purport to know enough about applied ethics to spot where the anti-natalist's arguments go awry. In general, there will be arguments for which one is not in a good enough position to make sufficiently fined-grained, accurate assessments, the kind one would expect from experts. It would be hubris to think that we could always do this. And, sometimes, the experts won't know either; it will be controversial.

Nevertheless, it's also hard to see why theoretical evidence should move us to make quite substantial changes to our belief systems. After all, we shouldn't overlook the important dissimilarities between our deepest convictions and our other beliefs. For starters, revising our deepest convictions would not be a mere change in *what we think* but in *who we are*. Changing one's belief that their children's lives have value, for example, is quite unlike changing their belief about whether the tip for lunch was calculated correctly. The changes in the former can affect one's values, one's sense of self, meaning, personal relationships, and identity (see van Leeuwen 2022). Why think that changing who you are should fall within the same normative constraints as your everyday beliefs? Perhaps theoretical evidence is not merely practically idle here, but normatively: it is permissible for us to treat it differently that other kinds of evidence.

Some will say that we shouldn't believe philosophical or other kinds of theoretical views anyway, so that there need be no pressure exerted from even the best revisionary arguments to reconsider our more ordinary beliefs. Instead, we can believe things like "My children's lives are meaningful" and yet *advocate* for nihilism about meaning in life without any incoherence. In effect, we champion these positions (see Goldberg 2013). The fact that these different domains command different kinds of attitudes means that the personal is effectively insulated from the theoretical. Call this the *Championing view*.

Another response is that what we think about in highly theoretical contexts and in ordinary life is simply different. Sure, we might say things like "I love my partner" or "my daughter's life is valuable", but what we mean by 'love' and 'value' is different than what we mean in the psychoanalyst's office or within the philosophy seminar. The objects of theory and ordinary belief are different, even if superficially they look the same. This is the *Context-dependency view*.

Others might say that there's something right about the thought that theory can bear on what we ought to believe, it's just that when we work with full beliefs, only then do we seriously risk having to substantially altering our doxastic economy when faced with exceptional theory. However, when we work with partial beliefs, this risk is greatly diminished. If I'm highly confident that my children have value (say, .95), but then an argument for a revisionary theory makes me less confident, I might only move a few decimal points and yet retain my belief that my children have value. This is the *Credence view*.

Lastly, some might think that theoretical evidence is typically permissive. The total theoretical evidence bearing on nihilism might rationalize confidence in nihilism, but it might equally rationalize suspending judgment. Indeed, your total evidence (including theoretical and non-theoretical) might rationalize belief *and* equally disbelief. So, when faced with what strikes you as compelling theoretical grounds for revisionary views like nihilism, skepticism, or anti-natalism, it might be rational for you to believe them, but you might also be permitted to suspend or even reject them. Call this the *Permissive view*.

I think these reactions are mistaken. The personal is insulated from revisionary theory for other reasons. Roughly, in cases where personal beliefs are at stake, we are permitted to treat theoretical evidence differently than other kinds of evidence because of what it would mean for ourselves and our relationships to integrate them within our total evidence. Insofar as epistemic norms council otherwise, they thereby tell us to divorce the belief from the believers in these cases, but that's precisely what we should resist. We would betray not only who we are, but the agents that fall within the scope of our evidential accounting, effectively treating them like exchangeable commodities in the marketplace of ideas. Call this the *Personal view*, or personalism.

Here, I focus on personal beliefs which are tied to intimate relationships. But others might wish to broaden the scope. Some will say that it holds of intrapersonal beliefs, such as beliefs that are central to one's individual identity.[5] Similarly, some might think that certain beliefs are necessary for accomplishing long term projects, which can justify more resistance to counter-evidence than garden variety beliefs (Morton & Paul 2018). Moreover, we might broaden the scope of who is wronged. Perhaps believing anti-natalism wrongs future humanity; or perhaps believing skepticism wrongs others with whom one bears no intimate relation. Although I think these ideas are plausible, I test the waters here with the more restricted version of the view. More specifically, I argue that we may bracket theoretical evidence and thereby permissibly fail to use our total 'theory-tainted' evidence to deliberate about what to believe. This enables there to be cases in which one is praiseworthy for epistemic irrationality. Importantly, practical reasons are not impinging on what we

[5] See Kawall (2010), who argues that enduring beliefs are necessary for autonomy, authenticity, self-control, and practical identity.

ought to believe directly; rather, practical reasons affect which bodies of evidence we may or may not consider.

This paper is organized as follows. §2 motivates personalism. §3 considers three alternatives: the context-dependency view, the credence view, and championing view, finding faults with each one. §4 explores how Mooreans would think of the epistemic role of theoretical evidence for personal convictions. Although we share the same ethos, ultimately Mooreans underappreciate the value of personal convictions. §5 further develops and defends the personalism before closing with what personalism suggests for practicing theoreticians, like academic philosophers.

2. Motivating Personalism

Consider the following case:

> **Influential Philosophy Paper:** Sarah is in a committed relationship with Clara and believes that [A] 'Clara loves me' entirely for personal reasons, like her experiences with Clara as well as Clara's testimony, actions, and so forth. One day, however, she reads in *The Journal of Prestigious Philosophy* an argument for the conclusion that love doesn't exist, and she can't identify any fault in the argument. It seems valid, and the author's argumentation for the premises look impeccable. As a result, she starts to feel unsure about A, thereby failing to believe that A. Still, she feels guilty about it. Luckily for her, a few days later she discovers a reply in *The Prestigious Philosophical Review*. She reads a very clear and compelling (apparent) rebuttal (unbeknownst to her and the author, it's misleading). This fully restores her confidence in A.

Put yourself in the shoes of Sarah's partner for a moment. We can appreciate what's so alarming about Influential Philosophy Paper by thinking about which conditionals would be true of you and why they would be problematic. If you were Sarah's partner, it would be right for you to feel wronged by her doxastic changes. The ease with which she altered her belief—from no longer believing that she loves you, to believing it again was double-mistake: she was moved by purely theoretical arguments twice over. Intuitively, as far as Sarah's partner can see, she shouldn't have been moved by those kinds of considerations in the first place.

Therefore, what cases like Influential Philosophy Paper bring out is that the following conditionals would be true of the rational agent:

(CE) 'If I hadn't read Philosophy Journal Article, I would still believe that I love you'.

Alternatively:

(CE*) 'If I had read Philosophy Journal Article, I would no longer believe that I love you'.

The intuition is that if these conditionals were true of you, you would be bad for it, lacking the right sort of concern for your relationship. Respecting the Socratic norm, to follow the argument wherever it leads, had a morally serious consequence: that Sarah should no longer believe that she loves her wife. What's problematic is not simply that she followed the argument where it led, but

that she did it twice. Her confidence was lost and restored only because of the arguments. It was only those theoretical considerations which led her away from belief and back again. *This* is what's objectionable: it represents an improper receptivity toward that evidence; she was far too concerned with being rational. The troubling thought, however, is that changing belief in response to her evidence is precisely what she ought to do; it is what rational thinkers would do. Certainly, ignoring that evidence is not an option for the ideally rational agent.

The personalist says that what's wrong with Sarah is that she's effectively *too Socratic*. More generally, she cares too much about following the evidence where it leads. She's too invested in fulfilling epistemic ideals. This claim should remind us of Susan Wolf's views about the moral saint, who's commitment to the right and the moral good leads to a deeply unattractive life, effectively spoiling their ability to live well. Fitting our actions so as to meet epistemic ideals likewise threatens to make our lives deeply unattractive, and even morally criticizeable. Just as reflection on the moral saint calls into question "the assumption that it is always better to be morally better", so too reflection on the epistemic saint, the person who aspires to meet all the demands of epistemic rationality as the Socrates within us is wont to do, calls into question the assumption that it is always better to be epistemically better (Wolf 2014, 28).

Here, I want to address several preliminary points. Addressing them will make clear personalism's commitments and help distinguish it from superficially similar, implausible views. The first point is that personalism does not permit bad believers—racists, xenophobes, sexists—to believe bad things. Revising bad beliefs sometimes threatens to unravel one's life, so why wouldn't this be sufficient reason for bad-believers to bracket contrary evidence? Even if the bad believer manifests integrity or respect for their intimates by failing to revise their bad beliefs in certain circumstances, say, because it would mean unknotting their loyalties to a hateful community, this just means that their commitments are already tied to criticizeable causes, goals, and people. While we can, in a certain frame of mind at least, praise the form of their concern—there's something right about standing for one's beliefs, honoring commitments, and so forth—we should criticize their content and practical consequences. Hence, even if we suppose that it would make the bad believer radically reorganize their thinking or ruin their intimate relationships to revise their bad beliefs, this is a cost worth incurring because of just how bad it is to harbor those kinds of beliefs in the first place. The personalist can remain neutral about *what* is bad-making about bad beliefs; it might be upstream or downstream considerations, or features of the doxastic attitudes themselves.[6] The point

[6] See Basu (2018) for the view that beliefs can wrong on their own, independently of upstream and downstream factors.

is just that, whatever makes them bad, it is independently strong enough to outweigh the prudential benefits of belief perseverance.

The second point is that the personalist's key normative claim is about the usability of theoretical evidence within our doxastic economy. The permission to retain personal convictions is conditional on what kinds of pressures the believer faces; it's a permission to look past theoretical evidence, evidence characteristic of philosophy, social theory, political theory, and so forth, but not any evidence whatsoever. At worst, personalism would permit bad believers to ignore liberal or egalitarian theory, for instance, but not the testimony of the oppressed about their experiences, their beliefs, their personal histories, and so forth, since these considerations are *not* theoretical.

Likewise, obviously a case in which one's partner testifies "No, I no longer love you" is evidence one ought not ignore, even though ignoring it might serve one's interest, given just how well it serves them to believe "Yes, my partner loves me". Crucially, this evidence is also non-theoretical; it doesn't require expertise, special training, domain-specific scholarly knowledge, or other academic skills to competently evaluate.

The third point is that we are theorizing what we ought to believe full stop and not what's epistemically rational. Here, one might worry that the puzzle thereby invites problematic comparisons between epistemic and pragmatic reasons. If one is presented with strong philosophical arguments against their personal belief, but it would be prudentially bad in the envisaged ways for the person to revise their belief, wouldn't the agent be tasked with comparing the relative strengths of the theoretical—and, so understood, *epistemic* reasons—with pragmatic reasons? This worry is premature. We can steer clear of bringing out the normativity scales to weigh the various reasons. For starters, when we are reflecting on cases like Sarah's, we are considering what she 'just plain ought' to do—what action she *really* ought to take—and we might think that epistemic reasons aren't the sorts of reasons with enough normative oomph to tell us what we just plain ought to do anyway; they're not authoritative reasons.[7] If that's right, there are no reason-kind comparisons in Sarah's case and so no weighing to make.

That said, even if we were forced into the more generic comparison of the weight of epistemic rationality against the weight of morality and eudaimonia, and fall on the side of the latter in these cases, one can still be properly criticized along one dimension of normative evaluation anyway: one is epistemically criticizeable—epistemically irresponsible, irrational, unjustified. This gives some rope to critics. But personalists should reply that sometimes we just shouldn't be that moved by charges of epistemic irrationality; sometimes, one shouldn't care *that much* about being epistemically rational.

[7] For this way of thinking about epistemic normativity, see Maguire & Woods (2020).

Cases like Influential Philosophy Paper raise a more general puzzle for how we should react to theoretical challenges to our personal beliefs. By 'theoretical challenges', I mean challenges from philosophy, theoretical psychology, political theory, sociological theory, and so forth. They are the sorts of challenges one might expect to learn about from an academic book, article, or lecture. They can be theoretical arguments, but also thought-experiments, theoretical interpretations of data, or higher-order evidence about expert theorist's attitudes.

Here, I want to further delineate kinds of theoretical challenges: theoretical challenges to ordinary widespread beliefs; to other theoretical beliefs; and to 'mixed' beliefs, beliefs which imbue a certain degree of theory.

For example, consider the challenge from physics to the belief that colors are visible properties of surfaces. This is a *theoretical challenge* in my sense, since it draws on a theory about the relationship between subatomic particles and colors to challenge the ordinary belief that colors are visible properties of surfaces. Many hold this belief as a kind of natural, unreflective judgment about the world.

Next are theoretical challenges to theoretical beliefs, beliefs like 'properties are abstract', 'Marxism is true', or 'We should read Beowulf from a structuralist point of view'. Theoretical challenges are then primarily argumentative, drawing on deductions, thought-experiments, intuitions, and theoretical considerations, coherence, simplicity, and explanatory power, drawing on theoretical concepts, definitions, distinctions, and so forth. What makes them *theoretical* is not one's degree of confidence in the positions, methodologies, or definitions but the disciplinary setting and methodology; as the definition suggests, *x* is 'theoretical' partly when it's not based on experience or practice.[8]

But there are mixed-cases as well. That animals should not be eaten is a philosophical position, but it can be part of one's religious creed, and deeply personal. Even more vague include 'the Earth has existed for many millions of years'—which are disputed by some fundamentalists and conspiracy theorists—but mostly widely accepted. Over time, theoretical views become entrenched; "fluid propositions" harden, as Wittgenstein put it (1969, §96).

Some theoretical challenges, however, contest obviously *non-theoretical* beliefs. This is what our puzzle is all about. I'll first delineate theoretical from non-theoretical beliefs by way of intuitive examples, before drawing out their different roles in practical agency:

Theoretical	**Non-theoretical**
It's not the case that anyone has a meaningful life.	Your children have meaningful lives.

[8] See Oxford English Dictionary, 'theoretical': https://languages.oup.com/google-dictionary-en/

Love does not exist.	You love your spouse.
There are no composite material objects.	Your best friend made you a dining table as a wedding gift.
Only one thing exists.	You have you nice friends.
No one knows whether there are external things.	You know that you have children.

While this distinction is relatively clear, it's not watertight. There are still other vague cases, like 'God exists' and 'We are free'—they are theoretical, but likewise widespread and sometimes personal. Here, I'll focus on the latter. Joan of Arc died for her faith; some, like Camus, counsel us to "rebel" against unfreedom and meaningless, understood metaphysically. What's going on in these cases? Here's one view: these beliefs can reflect commitments of agents rather than only a stake one has in a debate. What this supports is the idea that the same belief type— "God exists", "We are free", etc.—can play different practical roles for different agents.

To remedy this, it's helpful to distinguish between a theoretical from a non-theoretical belief in terms of the practical and affective profile of believing *p* for a person, and the role the attitude plays in the person's orientation to their own life. When *p* is personal, belief revision has *personal costs*, costs which might alter their conception of self, their relationships, or their worldview. It's not simply that there's a risk of reorganizing one's belief system, but that the *kind* of reorganization belief revision risks are serious changes for one's sense of self, their interpersonal relationships, and their practical commitments. It's one thing to give up on desire-satisfaction theories of well-being after so many years of defending it, and quite another to give up on believing that your children lead worthwhile lives, or that all this time your love was illusory. By contrast, theoretical beliefs are often used in theoretical reasoning, whereas non-theoretical are often used in practical reasoning. Although we might care about whether *p* regardless of its source, personal beliefs are often deeply aligned with the self—who one is or takes oneself to be—and interpersonal relationships—who one cares about—as well as one's commitments (whether to a cause, a project, a person, etc.) While this distinction is imprecise, it's sufficient for our purposes.

Here, I speak of *belief* rather than *faith* for two reasons. One: Although there are similarities, faith suggests a "weak epistemic position vis-à-vis the proposition in question", but our potential epistemic positions with respect to propositions like "I have good friends", "My children's lives are valuable", or "I love my partner" seem strong enough to have more than only faith in their truth (Alston 1996, 12).

Nevertheless, personal beliefs exhibit at least two faith-like qualities:

(1) They seem to be 'hot' cognitions: our personal beliefs have an emotional, affective dimension. One feels secure that *p*; one would feel lost if *p* were no longer treated as true, one is content that *p*, and so forth. These beliefs seem even more integrated with our individual and social identities than our other beliefs. Belief is part of social identity (Aguiar & de Francisco 2009; Van Leeuwen forthcoming). Empirical studies support the idea that qualify revision of certain beliefs as changes to who we are (Heiphetz, Strohminger, & Young 2016). Personal beliefs are integrated with our practical commitments; in what practical projects we undertake, how we organize our time, how we understand ourselves and our values, who we associate with and so forth.

(2) They are less responsive to counter-evidence than everyday beliefs, like beliefs about the weather, time, etc. The experience of being presented with counter-evidence to (DEPARTURE) "the train leaves at platform 5" is not like (MEANING) "My children's lives are meaningful". As soon as we hear the announcement that the train will now leave at platform 13, our commitment to (DEPARTURE) is lost. Even if the stakes are quite high—perhaps it is really important that we catch the train—this will lead us to double-check the announcement, but not belief perseverance if we get more evidence that the train departs from platform 13. That's not what happens with (MEANING); we will persevere. Here, I don't theorize about why this is or whether it renders these beliefs irrational. The point is just that the phenomenal experience of belief upkeep is more faith-like than everyday belief.

Two: if we were to bring personal beliefs under some other attitude category aside from belief, we risk "depopulating the category of belief" (McCormick 2023, 13). However, there's no reason from psychology to make this move. The fact that personal beliefs are hot cognitions and more evidentially resistant than garden variety belief doesn't warrant treating them as cases of faith, hope, or some other attitude. As Van Leeuwen & Lombrozo (2023) emphasize, "there's no guarantee that believing will be one thing" (2023, 3). We shouldn't expect the fact that personal beliefs don't pattern with perceptual beliefs that they're thereby something else. Indeed, some philosophers hold that even truth is not a constitutive norm of all beliefs (Hannon & DeRidder 2021). Concepts like 'identity-related beliefs', 'meaning-making beliefs', and so forth are common currency in cognitive and social psychology. They acknowledge the role of emotional, prudential, and meaning related concerns motivating belief (see Gelpi et al., 2020). One major idea here is to treat belief as a multidimensional type. Some function to track reality accurately, while others function to signal group membership (DeCruz 2020; Kahan 2015) or specific group commitments (Funkhouser 2020; Funkhouser 2017; Williams 2020). Some perform interpersonal and existential functions, like reducing anxiety and facilitating meaning (Inzlicht et al., 2011). And studies of ordinary language processing suggest that we use 'belief' to refer to stronger, stable convictions (like religious beliefs) than everyday belief (Heiphetz et al. 2021; Van et al. 2021). What we likely

ordinarily mean by 'belief' are those doxastic attitudes that embody both practical, existential, and interpersonal commitments *and* affirmations of what's true or false.

Now, when it comes to theoretical challenges, we often think that we shouldn't be like many of Socrates' interlocutors from Plato's dialogues. Socrates raised serious objections to their quite ordinary beliefs about things, drawing out their problematic consequences. Some of his interlocuters wanted to ignore them; doubtless some did. We think this reaction is blameworthy; it reflects a kind of *theoretical closed-mindedness*, a tendency to not consider seriously any theoretical challenges to what one believes.[9] On the one hand, then, it seems like the following is plausible:

> THEORETICAL SENSITIVITY: We should consider seriously the relevant theoretical challenges to our beliefs once apprised of them; i.e., the theoretical objections—once raised, from reasonable sources—are prima facie not properly ignored.

The trouble is that Theoretical Sensitivity can lead us to alter our deeply personal beliefs, as we saw in Sarah's case. But then can't we easily be blameworthy for so dramatically changing what we believe: our revision risks our authenticity, integrity, and easily makes us look heartless. This suggests that we also aspire to respect the following ideal:

> PERSONAL FIDELITY: One is prima facie permitted to not alter their personal beliefs just because of theoretical considerations, i.e., theoretical controversy, arguments, or objections—even if appropriately raised, from reasonable sources—are sometimes permissibly excluded.

Fidelity to our convictions can be a good thing. We praise people for their persistence in the face of argumentative adversity when their beliefs reflect "a normatively special and significant relationship a person bears to her projects and commitments" (Scherkoske 2012). Moreover, beliefs like "I love my children", "they're lives are valuable", and so forth reflect commitments we wholeheartedly endorse. We don't merely accept them. We *endorse* them. If you were to revise those kinds of beliefs in response to theoretical evidence alone, the charge of wishy-washiness would be apt since "integrity involves fidelity to one's endorsements" and this precisely the kind of situation where one should manifest integrity; theory is a right reason to alter one's research projects but not who one is (Calhoun 1995).

But fidelity can be highly problematic. It can make us worse. The problem, then, is finding the sweet spot where we manifest the right amount of theoretical sensitivity and the right amount of integrity and fidelity to our convictions. For respecting both norms maximally looks impossible; we

[9] Battaly (2018) thinks of closed-mindedness in terms of not considering relevant alternatives seriously. This conception is applied here.

must sometimes compromise one or the other. I'll first explore options which try to reconcile them (§3), but later I'll argue that we should favor personal fidelity, conceding theoretical sensitivity.

3. Reactions

This section considers three initial reactions to the puzzle and argues that while we can learn something important from each of them, ultimately they are insufficient. Later, we'll consider how Mooreans might respond before defending an alternative.

3.1 Context-dependency

What, exactly, is the connection between theoretical propositions and more ordinary propositions? Consider Bob, who believes that his children's lives are meaningful.[10] Perhaps you think that although Bob believes that his children's lives are meaningful, his belief expresses one proposition in the seminar room, and another when he's taking his children to the park and witnesses their enjoyment. The key idea here is that Bob really has two consistent beliefs, the belief (B1) "that my children's lives are meaningful$_{O}$" and (B2) "that my children's lives are meaningless$_{PH}$", whereby what individuates them are the fact that different meaning concepts are employed: meaningfulness$_{O}$, that is, ordinary meaningfulness, or *narrow meaningfulness*, and meaningfulness$_{PH}$, that is, meaningfulness in some broader sense, or *broad meaningfulness*.

After all, there is some evidence for this. When Bob is at the playground, his judgment that his children's lives are meaningful seems to be sensitive to different norms, particularly norms about how healthy they are, how happy they are, or how well their lives are going as judged by their parents. That is, his belief is sensitive to norms about narrow meaningfulness. And that's why he judges (perhaps correctly) that his children's lives are meaningful, because they satisfy the relevant narrow norms. When he is in the seminar, however, his judgment that his children's lives are meaningless seems sensitive to other factors. For example, these factors include whether their lives are purposeful from the point-of-view of the universe; whether their lives are likely to make a lasting, historical impression; or whether there is any objectively good reason for them to exist. And so on. That is, his belief is sensitive to norms about broad meaningfulness. And that's why he judges (perhaps correctly) that no one's life is meaningful, and hence not his children's either, because they don't satisfy the relevant broad norms.

[10] This example is based on a conversation with an expert in another discipline, reflecting on their beliefs before and after having children. He said he was a nihilist but then, in a transformative moment, lost their conviction as his daughter was born.

Although this just one case, it easily generalizes. The basic idea is that many of our personal beliefs which seem inconsistent with various philosophical positions are only superficially in tension with each other; semantically they are rationally co-tenable.

While this picture of the relationship between the theoretical and the personal is attractive, it is explanatorily limited. This is because people often experience a tension between what they discover (or think they discover) in the seminar room and what they believed before (and after) they arrived in it. This experience of tension after their theoretical discovery is some evidence of inconsistency between their earlier and later attitudes. It has a certain phenomenology: a phenomenology of surprise, distress and worry, or even relief, which at least seems to indicate that there *is* a conflict between his prior and later attitudes. Put differently, it's some prima facie justification to think that there is conflict between the two attitudes.

For example, perhaps Bob cannot identify any misstep in the one of the arguments for nihilism. This might distress him, and not because it is like an unsolved logic puzzle—it is *distress* rather than mere annoyance or frustration—and one way of explaining why is that Bob can see the potential implications of his new attitude: that no one's life means anything at all; not his, not his spouse's, not his children, grandchildren, and so on.

If, on the other hand, we supposed that Bob all along believed that his children's lives were meaningful$_{O}$, along with many other people's lives, and then discovers that no one's life is meaningful$_{PH}$, including his children's, his distress would be coherent only if he already believed prior to his discovery that his children's lives are also meaningful$_{PH}$. But we have no reason to suppose this; it's not at all necessary to make sense of Bob's prior judgments about his children or his actions with them, for example.

A second explanatory limitation with this idea is that it systematically threatens the significance of theoretical inquiry. If, generally, the concepts we employ in our personal beliefs are distinct from the concepts we employ in our theoretical beliefs—say, 'chair' means another thing in the philosophy room than it does in the living room, 'love' means another thing when you think 'I love my spouse' than when a psychoanalyst thinks 'people only love themselves', or that 'freedom' means one thing for the protestor, and quite another thing for the metaphysician—it will be difficult, if not impossible, for theoretical inquiry to improve, revise, or be informed by facts about non-theoretical belief and the concepts they employ.[11] That's because these domains aren't using the same concepts. Discoveries about 'chairs' in metaphysics, 'love' in psychoanalytic theory, or 'freedom' in political theory, just won't tend to be discoveries about the stuff we think about in our

[11] I've encountered this idea in conversation. For something like it in the case of ordinary object terms, see van Inwagen (2014).

living rooms, bedrooms, or picket lines. Socrates' hope of helping ordinary people reorganize their thinking would be completely confused.

3.2 Credence

One might think that adjustments in confidence above .51 don't necessarily compromise one's integrity or seem to be as ethically problematic as abandoning belief altogether. For example, we can imagine that if, after having appreciated the argument from Influential Philosophy Paper, Sarah had become slightly less confident that she loves her partner, this would not have compromised her integrity, nor would it have been disrespectful towards her partner.

I think we should be sympathetic with this idea. Obviously, *some* adjustment in confidence should be permissible. Perhaps no doxastic attitude—no matter its morally good-making features—requires such exactness of commitment. Rather, it just requires some threshold sufficient for genuine commitment.

But this idea only goes so far. Consider Bob again, but this time he adjusts his confidence as follows:

> Bob: "I used to believe that you all have value, but now I'm only slightly more confident than not that your lives have any value."

Here, Bob quite clearly wrongs his children. His mere confidence is hurtful. He should *believe* that his children have meaningful lives. It's hurtful for him to be, say, around .6 confident that this is true, as if he's placing a bet. Likewise, consider Sarah prior to reading the reply:

> Sarah: "I used to believe that I love you, but now I'm not as sure; *I'm just more confident than not."*

Here too, while Sarah is not unconfident that she loves her partner, she is only more confident than not that she loves her, i.e., some decimal point just greater than .5. What these cases reveal is that the threshold for commitment—or, at least, commitment that is not prima facie morally impermissible in a case—is rather high. Slight changes are morally permissible—perhaps the move from full belief to being just below it is too slight to matter—but *mere* confidence is not. This points to the idea that personal commitment, understood as partly a manifestation of the agent's doxastic attitudes, is a rather strong attitude, greater than mere confidence but perhaps not as great as certainty.

Hence, employing credence doesn't avoid the puzzle. Indeed, it might even strengthen it. Consider Influential Philosophy Paper again. Sarah can't identify a flaw in the argument, and if she's honest with herself, she's not sure there is a flaw. Nevertheless, she doesn't believe the argument supports the view, nor does she judge that it is likely sound. She's just not sure. So, her appreciation of the argument leads us to become a little less confident than she was before about whether love exists and, in turn, whether it's true that she loves her partner.

The problem is that someone who displays that kind of sensitivity to theoretical argument might strike an intimate as a character flaw when its bears on questions that involve them personally. It's not that one shouldn't care about evidence, or that it is wrong for one to be persuaded, if only ever so slightly, by theoretical considerations, but that one displays an unkind sensitivity to theory when they are willing to alter their interpersonal beliefs in light of theoretical considerations alone. When Clara learns that Sarah *actually doubted* that she loves her because of a theoretical argument, we might say that she is entitled to expect a special explanation and restitution; something that softens the blow, and eventually repairs the wound. Still, the sense of harm is justified. It is something we may reasonably expect of an intimate. Crucially, the intimate's experience is directed at the doubting agent's credal state, or more specifically their transition from conviction to a weaker confidence state. If that's right, then the strategy of dissolving the puzzle by replacing belief with credence will likely fail.

A related strategy is to urge that people like Sarah should retain their beliefs but adjust their credences, understood as the realization of distinct mental state types.[12] So, Sarah might continue *fully believing* that 'I love my partner' but have much lower confidence in it. In this way, she could consistently retain her conviction and yet doxastically appreciate the weight of theory.

Unfortunately, this possibility is not easily available to the reflective agent. Imagine Sarah affirms her pair of attitudes: (BC1) "I believe my partner loves me, but I'm not so confident that she does". That looks Moore-paradoxical. Consider also: (BC2) "I believe my partner loves me, but I wouldn't bet on her loving me", where the latter reflects her more precise subjective probability judgment. BC2 certainly seems hurtful. Imagine Clara overheard Sarah expressing her attitudes. She'd have reason to think Sarah is deeply conflicted, and she'd be right. In turn, introducing credal states to resolve the puzzle is neither a moral nor a rational victory.

3.3 Championing

Consider the person who fully believes that p, but actively champions that $\sim p$ professionally. For example, imagine Jeff, who *believes* "my child should have been born", but who *champions* anti-natalism, the thesis that it would have been better if no one had been born, where this entails that 'no one ought to be born'. In turn, he qualifies in the profession as as anti-natalist.

A first glance, we might think Jeff is incoherent, and, indeed, hypocritical. He's incoherent because, by believing that:

(SB) "My child should have been born".

[12] This is called belief-credence dualism, on which there are beliefs and credences, and neither reduces to the other. See Jackson (2019).

He manifests his commitment to it; he takes a decisive stand on the question about whether his child should exist.

However, by championing anti-natalism as a professional academic philosopher, he actively argues for anti-natalism, the thesis that no one should be (or have been) born, which trivially entails:

(~SB) "It's not the case that my child should have been born".

If he were to believe ~SB as well, his beliefs would be rationally incoherent: he would be committed to *p* and ~*p*.

But Jeff escapes the charge of rational incoherence, since he only *champions* anti-natalism. What this means, minimally, is that (i) he actively argues for it; (ii) he defends it from objections; and (iii) he is willing to use it to fulfill certain (domain-specific) practical ends, like publishing articles, giving lectures, holding debates, submitting research grants, reviewing manuscripts, and so forth.

Moreover, and more to the point of the puzzle, Jeff seemingly escapes the charge of theoretical sensitivity. It's not like the philosophical arguments changed any of his *beliefs*, only his professional behavior, like what arguments he mounts and defends, what grants he writes, what lectures he gives, and so forth.[13]

One argument for championing our philosophical views is that we can do everything we could do professionally by championing them that we could do with believing them. So, we might think that, on balance, championing positions is less epistemically risky than believing positions, since we might be unjustified or irrational to believe whilst justified or rational in championing.

I want to grant that this is right but argue that it doesn't fully resolve the puzzle. We can appreciate this by revisiting the case and changing it in two ways so that the protagonist is, first, (i) a non-practitioner, and so only a consume*r*, and second, (ii) an expert, and so a practitioner or producer. In the consumer case, I argue that championing is not a live option; it's interpersonally confusing because it misleadingly signals conviction. In the expert case, I argue that, although championing doesn't involve any rational incoherence, it is often still hurtful.

Suppose you champion anti-natalism. You publish papers defending the view. You give lectures arguing for it, rebutting objections, and developing its applications, like future projects that align with it. You've built your career around the defensibility of the view. However, you also (of course) believe that your children should have been born. In the hospital, for example, at your children's birth, this belief had the status of a certainty—doubting it was beyond the pale. How could people make sense of this combination of attitudes and activities? Consider 'clarifying positions', which are statements which the agent uses to clarify their attitude and action profile:

Clarifying positions:

[13] Of course, this is consistent with the theoretical arguments altering his adjacent beliefs, like about which arguments are plausible, which theories have a better chance at being right than others, and so on.

(1) "Don't worry Clara, I do believe that you love me. I just seriously defend the view that no one loves anyone; I rebut the critics, I educate the students, I try to convince the grant committees, among many other fitting actions, but I don't defend the opposite or any contrary opinions.".

(2) "Don't worry James, I do believe that you should have been born. I just seriously defend the view that no one should have been born; I rebut the critics, I educate the students, I try to convince the grant committees, among other activities, but I don't defend the opposite or any contrary opinions".

Clarifying positions are troubling because they suggest a lack of integrity on the part of the clarifier, or even hypocrisy.

To foreground this idea, consider paradigm cases of integrity deficits. When we think of the committed anti-capitalist who nevertheless works in and excels at financial speculation and venture capital investing, we see them as lacking integrity; as contravening their values. Indeed, it makes us question their authenticity: what kind of committed anti-capitalist actively excels in a career in high finance? Irrespective of what we think of the person's values or their reasons for valuing what they do, they are still independently subject to criticism for their hypocrisy.

When it comes to commitments about the worth, value, or even the existence of our loved ones, belief represents a settled attitude. If you believe that your children's lives have value, that their living is worthwhile, it would be odd to petition a protest against other people having children, present self-help talks to council families against biological parenting, to create a meetup for the advocacy of anti-natalism, and so forth. You look like a hypocrite. Plausibly, you would be.

Now, imagine all the while you are contentedly pregnant, or a supportive partner and parent. Here, the charge of hypocrisy easily sticks. "What are they doing?" is an appropriate question. You would look phony, and at best, confused.

In this case, we imagine someone who championed that p but believed that $\sim p$, and yet participated in the canonical actions, befitting the content or practices surrounding that content, associated with belief or other doxastic commitment. If we think there is something criticizeable about the anti-natalist advocacy case above, why shouldn't we think the same about the philosophical advocacy case?

Another issue with this idea is that it's rationally unstable. The general motivation behind championing theoretical positions is that we just don't or even can't justifiably believe such positions, and so we should thereby withhold belief in them. Although I will draw on recent work on belief in philosophy here, the point naturally extends to the humanities and other areas of highly theoretical inquiry where there is widespread peer disagreement as well.

Goldberg (2013) argues that recognizing systematic peer disagreement about a philosophical position makes it unreasonable to believe it. This is because it would be 'normatively inappropriate' to believe it whilst believing that the balance of evidence doesn't favor it or its contraries, something we are committed to by recognizing systematic peer disagreement over P (2013, 9). As a philosopher, one should recognize that, for any proposition *p* for which there is systematic peer disagreement, the arguments for and against *p* are not decisive. To the question: "which side is better supported by the evidence?" one's attitude should not have the level of confidence that pairs with belief, which is what speaks for not believing one's position or its negation (see Goldberg 2013, 11). Otherwise, one would be in the Moore-paradoxical position of accepting 'P, but I'm no more confident that the evidence favors P than its negation'. While some theorists say that there are cases of rational 'level splitting', it represents a position one is forced into rather than one that thinkers seek out; it would be better if we didn't ever have to endorse both '*p*' and 'my evidence doesn't favor *p*'.[14]

Still, there is a sense in which one can 'have a view' in theoretical debates whilst lacking belief. Goldberg says that having a view is a matter of *regarding the position as defensible*, whereby one is committed to defending it when the situation arises. Although Goldberg's thesis about regarding-as-defensible is controversial, I'll grant it here. What I take issue with is whether it is the strongest permissible attitude we can take in cases of theoretical controversy, given plausible closure principles that we have few if any independent reasons to deny. Here's the gist of the argument. Suppose it is rational for us to believe claims like (MV) "My children's lives are valuable". MV entails the negation of several theoretical positions: value nihilism, other minds skepticism, external world skepticism, and solipsism, for example. Mutatis mutandis for similar personal beliefs; they'll logically entail the denials of many other theoretical positions. You can competently deduce from MV that certain theoretical positions, like nihilism about value, is false.[15] If we should only regard-as-defensible our theoretical positions, however, we cannot rationally come to believe those positions on the basis of rationally believing our personal convictions and our rational recognition of their logical implications. Hence, either closure for rational belief is false, we don't rationally believe ordinary personal claims like "my children's lives are valuable", or else Goldberg is wrong that we may only rationally regard-as-defensible highly theoretical positions. I submit that Goldberg's position is the weakest link here.

Here is the argument spelled out more fully. Consider a simple argument, like:

[14] For defenses of level-splitting, see, e.g., Lasonen-Aarnio (2014).

[15] One might argue that disagreement about value nihilism is not widespread enough for it to be a sufficient higher-order defeater for one's belief in a logical contrary of it. As a proxy, 14.90% of philosophers surveyed accept or lean towards the non-existence of meaning in life, and 4.17% are agnostic. However, only 5.27% accept or lean towards error theory in meta-ethics. See: https://survey2020.philpeople.org/survey/results/all

Simple Argument:

(1) My children's lives are valuable.

(2) If my children's lives are valuable, then value nihilism is false.

Therefore,

(3) Value nihilism is false.

Suppose you believe (1) and (2). If you have children, then you probably believe (1). (If not, replace (1) accordingly: my best friend's life is valuable, my brother's life is valuable, etc.). Clearly, insofar as you have children, there's a sense in which you should believe that your children have valuable lives, or at the very least lives that are not worthless (perhaps bracketing special circumstances, like with a kid who spends all their time aimlessly smoking marijuana, or scrolling Instagram). Even more weakly, that they have a shot at living a valuable life. Value nihilism, however, says not merely that *in fact* people lead valueless lives, but because of the absence of value in reality, they just can't lead valuable lives. No one's life could have any value. On (2): setting aside contextualism, standard interpretations of (MV) "My children's lives have value" and (VN) "No one's life has any value" are logically inconsistent. VN says that there is no x such that x has value, or $\forall x \text{ value}(x) \rightarrow \neg\exists y$ value(y) and x=y, while MV clearly states that there are people (your children) who have value.

From modus ponens, then, you can draw the conclusion (3). But since you ought not believe (3), and yet you permissibly believe (1-2), you're in the odd scenario in which you rationally may believe the premises of a modus ponens argument but rationally ought *not* believe its conclusion. Instead, you may only regard it as defensible.

There's two points to make here. The first is that Goldberg's position seems to mandate a violation of a plausible closure principle, namely that if you rationally believe P, and rationally believe—indeed, know—that P implies Q, you are thereby in a position to rationally believe that Q. Since you rationally may believe that (1-2), but not (3), you should reject closure for epistemic rationality.

The second point is that, insofar as we are attracted to closure for epistemic rationality, we should say that we *are* rationally permitted after all to take an attitude stronger than regarding-as-defensible towards at least some positions subject to systematic peer disagreement, namely those positions which (a) we already believe, (b) enjoy some 'special doxastic status'—maybe they're common sense, maybe they're part of what Kelly (2008) calls the 'hard core' of common sense, or maybe they have a central meaning-making role for us, whereby revision would rationally require an overly demanding reorganization of our other beliefs and commitments—and (c) they are contraries of the target theoretical positions. Whatever the story, those attracted to the idea that we may take two different commitment attitudes, one to the revisionary theory, the other to the contrary personal conviction, have a steep hill to climb.

3.4 Permissivism

Permissivism says that the evidence permits one to believe that *p* but it also permits one to disbelieve that *p* or suspend judgment about whether *p*.[16] In this fashion, Permissivism makes space for the person who, in cases like Influential Philosophy Paper, resists following the argument where it seems to lead.

The idea is that the argument in Influential Philosophy Paper is a piece of (theoretical) evidence, *E1*, and that Sarah's experiences with and testimony from Clara is another piece of (personal) evidence *E2*, both of which bear on the question about whether Clara loves her. Putting them together with her evidence about love illusionism's (apparent) entailments, her total evidence *E* permits (B1) believing the conclusion of the argument, (B2) believing that the argument is unsound, or (B3) suspending judgment about love illusionism, and so, intuitively at least, suspending judgment about the relevant entailments. That is, her total evidence doesn't decisively favor any one of option, but permits any of them.

Permissivism, however, doesn't quite vindicate the core intuition in play. The intuition is not just that following the argument where it leads in Influential Philosophy Paper would be wrongful, but that integrating the evidence into one's total evidence, understood as the evidential base from which one deliberates, would be wrongful. In that case, the target of our evaluation is the *deliberation* and *weighing of evidence* as well, not only the doxastic change.[17] Modestly understood, factoring in the evidence from the philosophy article together with their reflection on the argument ought to lead one to reduce their confidence, and not just ever so slightly. Permissivism clearly presumes that such evidential integration occurs.

But what's especially worrying is not quite the evidential integration, but the permission to follow the argument where it (apparently) leads. Sure, one is also permitted to do otherwise, but strictly speaking if Permissivism were true, blaming Sarah for her doxastic transition would be improper and undeserved, since her doxastic transition would be epistemically rational—it's just one of the epistemically rational options available to her. For this reason, although Permissivism might improve upon on the other strategies, it doesn't account for the intuition that altering one's belief in cases like Influential Article would be wrong, bad, or blameworthy.

---

[16] Here, our focus is on intrapersonal permissivism, which is about the permissible attitudes for a single agent towards *p*, given a body of evidence *E*, rather than distinct agents towards the propositions. For a defense of intrapersonal permissivism, see Jackson (2021).

[17] Here, one might have worries akin to those for doxastic voluntarism. How could we fail to avoid weighing evidence and deliberating once the evidence finds in way into our cognition? One answer is that we shouldn't underestimate motivated cognition: of avoiding certain bodies of evidence because of their (perceived) conflict with what we believe. The fact that we have been made aware of the evidence antecedent to them: to the intention to deliberate, to spend time weighing evidence, to be careful in one's weighing, or generally that doesn't try to do things that would offset deliberation. Thanks to Jakob Olhurst for this suggestion.

4. Mooreanism

Mooreans think that our common-sense beliefs enjoy a secure epistemic status that philosophical argument does not remove (Kelly 2005; Fine 2001; Hirsch 2002). Here, we'll explore an extension of this position to personal beliefs. While I think personalism and Mooreanism are similar in ethos, we differ in important details.

Some Mooreans will argue that we tend to have stronger evidence for our personal convictions—"My children do meaningful things", "I love my partner", "I know I have good friends", etc.—than theorist's do for their contrary theoretical beliefs. This is why there is no serious pressure on people like Sarah to revise their personal convictions when faced with theoretical arguments. Yes, theory is evidence, but it's comparatively bad evidence.

Think for a moment about how you would support beliefs like "Of course I love my partner", "My children are valuable; I wouldn't trade them for the world", "Sure, I remember that I'm meeting my friends for brunch tomorrow". It can be difficult to put into words; our support for them comes from a dynamic interplay of experience, testimony, emotions, embodied engagement, intimacy, a 'common consent' amongst the people we routinely interact with, and so on. One might say: it's hardly the result of theory and interpretation.

By contrast, the evidence for revisionary theories is purely theoretical, nearly by definition: they do *not* align with common sense; they are often *at odds with* pre-theoretical intuitions; our everyday experiences do *not* confirm or even suggest that they might be true. The support for revisionary theories depends on other theories, on controversial interpretations, on theoretical arguments involving highly idealized or even science fiction scenarios, and different ways of weighing purely explanatory considerations.

Of course, as the Influential Philosophy Paper highlights, revisionary theories also depend on theoretical arguments, but the point is that these arguments depend inexorably on theory-building and interpretive considerations: on how to weigh intuitions about idealized thought-experiments; about whether, say, normative unity is important; about whether to understand human interactions through the lens of Aristotelianism, psychoanalytic theory, Marxist theory, and so on. A person attracted to, say, radical skepticism might think that its easy alignment with their intuition in the new evil demon case, content internalism, and epistemic internalism, is an explanatory good-making feature, given their prior acceptance of those other theories. Obviously, those considerations won't be shared by every theorist; some will see that theory alignment as a bad-making feature. Hence, we might think that, on balance, the evidence at issue in personal cases is stronger than the evidence at issue in the relevant theoretical cases, so that there's little if any normative pressure exerted from theory to one's personal convictions.

A related idea is that the person who should retain their prior belief is the one who has the *right reasons* for believing what they do in the first place. The right reason-makers are objective epistemic qualities, like factivity, objective probabilification, or alignment with objectively true epistemic norms. The application of the idea here is that maybe the person who believes that "love is real" in virtue of personal convictions like "I love my child" has the right reasons for their view. They have much stronger reasons to retain their prior belief, and with it the theoretical position it presupposes, than the theoretician who denies it.

I think we should be suspicious of this strategy. One reason is that it moves the core issue. Cases like Influential Philosophy Paper make us uncomfortable with the thought of following the evidence where it leads. Consider Sarah again. It's not that her initial evidence was poor *and* misleading. Rather, her appreciation of the theoretical argument made her see that her initial evidence just isn't *good enough* anymore, and not that it's not any good or that it was never sufficient. The strategy which says "yes, but people like Sarah often have the stronger body of evidence to begin with" are in effect saying that Sarah is just making a *rational mistake*; that there *must* be something wrong with the argument, so that her new, larger body of evidence is misleading.

For example, Kelly (2008) argues that it would be irrational not to rely on what one knows when engaging in philosophical inquiry, because what one knows is part of one's total evidence, and we shouldn't leave our evidence behind when we engage in philosophical inquiry. Expanded to our case, the point is that we shouldn't leave our total evidence behind when evaluating theoretical arguments. Sure, theory is evidence too, and so we can't ignore it, but so too we can't ignore what we already know.

This point raises two questions. First, is our prior knowledge retained when we are confronted with theoretical arguments contrary to our prior beliefs? The second question grants that we do retain our knowledge, but asks whether we would be justified in countenancing revisionary arguments. If I'm right, there's something objectionable about factoring in such arguments as evidence, and not merely with what one makes of one's total body evidence once the arguments are given proper accounting.

Before we address these questions, it will be instructive to work though the Influential Philosophy Paper case more fully to make some educated conjectures about the quality of our total evidence in similar cases. If we suppose that, prior to Sarah's awareness of the theoretical argument, *A*, she knew that (L1) "I love my spouse" and (L2) "my spouse loves me", but then (T1) *A* seems to her to be sound; (T2) she can't identify any flaw in *A*, (T3) the thought experiment of the paper seems compelling to her, and (T4) she grasps the alleged theoretical virtues of the theory, then her total evidence now includes L1-T4. Ex hypothesi, L1-L2 are true, and so T1-T2 are only suggestive, while T3-T4 are misleading. But does her total evidence then favor continuing to believe what she

does, so that she need not alter her prior beliefs? If we think that a body of evidence E supports p only if E makes it more objectively probable that p, then Sarah's total evidence supports her prior beliefs. But if we think that misleading evidence can genuinely support a belief, then the answer is not so clear, for Sarah might be rationally required to be less confident than she was before, even if that doesn't mean removing her belief entirely (i.e., she doesn't need to suspend judgment, but she shouldn't be, say, >.8 confident in her prior beliefs either). This should make us squeamish. There's something objectionable about merely being *more confident than not* that you love your spouse (or vice-versa) given only theoretical evidence (see §3.2). Her spouse could rightly blame her for falling short of (full) belief in such a situation. The Moorean, then, who argues that we tend to have the right reasons for our personal convictions is making a bet: that we won't tend to have to make downward rational changes away from our convictions.

The second point is that there is instability in the position that we tend to have stronger evidence for our personal convictions than theorist's do for their contrary theoretical beliefs. After all, there's nothing per se objectively epistemically better about personal, non-theoretical reasons than theoretical reasons, the kind we get from arguments in philosophy, theoretical psychology, social theory, and so forth. If that's right, then we might worry that the approach we're considering now needs to suppose that we often fall on the 'right' side of such disputes, but why think that? Here, I'll take up this question in turns.

First, consider the gold standard in philosophy: deductively valid arguments with only true premises. If a sound deductive argument A for p is your evidence for p, then your evidence doesn't leave open the possibility that you're wrong. What *could* improve on your evidence here? No amount of evidence could make your true belief any more likely to be true, since the probability that p is true on your evidence is 1.[18] Evidence from theory, can, then, be very strong, and it's otherwise difficult to make sense of the notion that we have 'better evidence' in the personal, non-theoretical cases than the theoretical cases without reifying personal beliefs (cf. Kelly 2005, 188). What factors could *make* the evidence in principle better that doesn't depend on positing highly controversial features, like infallibility, acquaintance, or certainty, as the regular determiners of justification for such beliefs?[19] So, there's no compelling reason to suppose that theoretical evidence *must* be weaker than non-theoretical reasons or evidence.

[18] At best, what could enhance the epistemic quality of your belief is more diverse evidence, so that if you are led away from one sound argument, another sound argument might still exert its influence. But then the original point applies here as well: imagine you have many sound arguments for your view. If you have a diversity of sound deductive arguments, it's exceedingly difficult to imagine what you could do to evidentially improve upon it. In turn, if your belief is based on a diversity of deductively valid and sound arguments, it more easily endures critically scrutiny.

[19] Of course, one might argue that theoretical arguments for revisionary conclusions are usually unsound, but we shouldn't have to resort to this pessimistic metaphilosophical view about theory's evidential strength generally only to explain why exercising dogmatism against theory can be praiseworthy. Plus, it misses the mark: the intuition seems to be broadly ethical: one is praiseworthy in a way that isn't reducible to how well one models the ideal epistemic agent.

Second, consider Kelly's (2005) view that Mooreans are simply those who hold that it is "the correct norms of belief revision" rather than any special norms which protect common sense (and herewith, personal convictions) from revisionary theory (Kelly 2005, 187). In particular, the norm of belief that effectively insulates common sense belief is:

> *More Reasonable*: In resolving conflicts among one's beliefs, one should always favor those beliefs that it is more reasonable for one to think are true given the totality of evidence and arguments to which one has been exposed (ibid 194)

What's most epistemically reasonable for us to believe is whatever is our total evidence supports, which includes our first- and higher-order evidence. In particular, the first-order evidence is not more epistemically significant than higher-order evidence: both equally bear on belief, even if in some cases one type should be weighted more heavily than another. In turn, what *More Reasonable* implies is just that, in resolving conflicts among one's belief, one should always favor whatever one's total evidence supports.

In cases like Influential Philosophy Paper, how likely is that the theoretical argument weighs more evidentially than the remaining higher- and first-order evidence of the person who denies the conclusion? For example, the fact that neither Sarah nor Clara's friends or family are suspicious of their love for each other surely is strong higher-order evidence. Moreover, Sarah has strong first-order evidence, namely Clara's testimony, and her first-hand experience with her overtime, all of which suggest that they are in a loving partnership. In turn, we might naturally think that cases like Sarah's are typically ones in which one's pre-theoretical evidence is strong enough that, when pooled with their theoretical evidence, still favors their prior belief.

Although following the total evidence will tend to align with Personalism in the sense that, when personal beliefs are threatened by theoretical evidence, one ought to retain their personal beliefs, *More Reasonable* is morally questionable for the agent in a way that Personalism is not. To see why, let's reimagine Influential Philosophy Paper where Sarah resists instead of making the transition, because she recalls that her total evidence simply points to keeping her doxastic attitude. Imagine she thinks like this:

> *Explication*: "I have some first-order evidence that suggests that my belief about Clara and I is false. Some of my higher-order evidence also suggest this, because some experts are confident in a conclusion that is incompatible with my first-order belief. However, I have stronger first- and higher-order evidence for my prior belief about Clara and I. My total evidence therefore suggests that it's most reasonable for me to retain my prior belief".

Insofar as we thought that Sarah's transition from believing the conclusion to moving back to her initial belief was heartless, so too we should think the same about *Explication*. The worry is that the agent *disinterestedly* follows the total evidence, and not quite that she respects the total evidence.

The Moorean might have a reply here, however. Contrast the virtuous with the vicious agent. Here, our focus is not on epistemic or moral virtue exclusively, but the wholly virtuous agent, one who exhibits excellence broadly and is generally admirable. Now, while virtuous agents will respect the total evidence, they will not do so always disinterestedly, such as when there are serious personal stakes.

When the evidence favors one's prior personal belief, one may be *relieved* or *assured*, and when it favors revising their prior personal beliefs, one may be *concerned* or *regretful*. That is, the virtuous agent couples her *intellectual* activity (following the total evidence) with the fitting *affective* activity (relief; content; sorrow; regret).

This is an improvement on the Moorean's basic reply. Still, their response can be unsettling. Without a general policy concerning the balance of the total evidence in favor of personal beliefs, sometimes the theoretical evidence and the activities surrounding it won't favor the prior personal belief. What turns out to be *most rational* is much more subject to luck than that, in the sense that it turns on the specificities of the agent's psychology, what evidence they are exposed to, the quality of that evidence, and so forth, all in keeping with the idea that—given those facts—they uniquely determine an attitude that is most rational for the agent at that time. In those cases where the balancing act doesn't favor the prior belief, the virtuous agent will need to regretfully reform not only what they think but who they are. Imagine what this would look like in cases where there are serious downstream practical consequences for friendship, love, and value:

> (M_V) "Look, I'm sorry but the evidence is too strong: regretfully, it seems like your life doesn't have any meaning at all".
>
> (L_V) "Look, it bereaves me to think this way, but the evidence regretfully suggests that your pronouncement of love for me is false, and our experiences with each other are likely misleading".
>
> (N_V) "Look, I'm really sad to have discovered this, but the evidence suggests that you shouldn't have been born. It would likely be better if you hadn't existed at all."

Even when we imagine them asserted with sincerity, it might still rub us the wrong way. Doesn't it reflect a deeper concern for *being rational* than for the wellbeing of one's intimates?

Moreover, if one is like the epistemic saint here—who's ultimate concern is with being epistemically rational, so that it's *not* just another value among other values—then we might take issue with their axiological favoritism. What kind of sacrifice is it *really* to risk doxastic irrationality? In these kinds of personal cases, isn't taking the 'risk' worth it? So, it seems like the praiseworthy response is for the agent to retain their prior belief but with higher-order hopefulness: the hope that one's belief is (or will become) epistemically rational. The Moorean's policy of giving personal beliefs

contra revisionary theory the benefit of the doubt is a good one, but not because it's always the most epistemically rational choice.

5. Personalism and the Exclusion Rule

What best explains our permission to resist theoretical evidence when it puts pressure on our personal beliefs, then? Context sensitivity, distinct attitude, credence, permissive rationality, better reasons, and total evidence views come with costs and otherwise fail to recover the core intuition that there's something wrong with revising one's personal beliefs in response to novel theoretical evidence alone. I now want to suggest that we are sometimes permitted to bracket theoretical evidence for moral and prudential reasons.

Consider the more generic idea that one can sometimes permissibly bracket evidence which bears on whether *p* and may thereby permissibly not make use of that evidence in deliberating about whether *p*. Call this the *exclusionary rule*. The exclusionary rule gains traction when we think about cases in which how we obtain or handle evidence matters to us for some practical end that we collectively care about.

For example, many criminal courts prohibit the use of illegally obtained evidence in judge or jury deliberation. Other kinds of evidence can also be inadmissible, like evidence through improper questioning or hearsay (e.g., using what someone else told a witness about the defendant as evidence).

Importantly, this prohibition does not (nor is it intended to) mean that the inadmissible evidence is poor, but that it is instrumentally defective, something that the agent in some designated role ought not use for the purposes of fulfilling their practical goal. In criminal cases, the thought is that illegally acquired evidence is not the *right kind* of evidence relative to the goal of reaching just verdicts.

Now, one might think that inadmissible evidence cases in criminal trials don't quite speak for the exclusionary rule when it guides doxastic deliberation, but something weaker, like the principle that if one has to decide *D* (like proposing the verdict 'guilty' or 'not guilty' to the court) about whether *p* (the defendant's guilt or innocence), then, if evidence *E* bearing on whether *p* has some instrumental defect (like being illegally obtained, relative to the goal of reaching legal verdicts), then one ought to exclude *E* for *D*. The difference is that inadmissible evidence cases speak for action or intention—making a decision—and not belief.

But certain kinds of inadmissible evidence cases seem to be proxy for what we might call "unjust" or "unfair" evidence cases. Below, I'll discuss two sorts of cases. Consider first a variation on the illegal evidence case:

Torture: The police tortured the defendant prior to their official questioning. Moreover, the defendant's testimony in a video recording, as well as the detective's expert testimony about the defendant's behavior while alone in the room prior, during, and after their torture, becomes privately available to you.

There is a legal prohibition on using the evidence at issue in the torture case to reach legal verdicts, and so reaching decisions, beyond that fact that torture is often unreliable vis-vi-vis the ends of acquiring accurate confessions. Intuitively, there seems to a *moral* prohibition on using that evidence to form certain kinds of beliefs as well.

For example, consider questions about the person's character—are they cowards, bad-tempered, shameless; ignorant or narrow-minded? Consider also questions about their intentions, motives, and criminal psychology: are they sociopathic, sadistic, or vengeful? Finally, consider questions about their intimate psychology: do they suffer from borderline personality disorder, suicidal ideation, or severe depression?

Making up one's mind about these sorts of questions on the basis of the kind of evidence made available to one in the torture case is intuitively improper. After all, believing that *p* is sufficient for making up one's mind about whether *p*; it reflects a decisive attitude about *p*. Although belief is not what we might call a "final verdict"—one might still inquire about whether *p* even after one has formed the belief that *p*—it is still a settled attitude; one takes a stand on whether or not *p*, even if one seeks to enrich their state by making it knowledge or more certain.

The point is that having settled attitudes about certain intimate questions, like those concerning the defendant's character, motives, or personal psychology, on the basis of the evidence gathered from their torture doesn't seem to treat them respectfully; in particular, it exploits their vulnerability in the situation, even if the evidence in fact raises the probability of certain answers to those questions. The thought is this: you would wrong *them* to use *that* evidence to form a settled attitude, like belief. Next, consider:

Secret: Your best friend Tessa doesn't seem to you to be doing so well, but she hasn't devolved any details about why this is, and you don't press the matter. However, her estranged sister tells you very intimate details about her—details which, if accurate, represent a very different picture of Tessa than the one you've developed over the years. They are details which you rationally believe Tessa would, if true, want to be kept secret.

Assuming that friendship imposes certain responsibilities on the friends, we can ask here whether friendship has special epistemic responsibilities, like not forming beliefs (as opposed to hunches, stray thoughts, etc.) about our friends on the basis of shared secrets. Call this supposed special epistemic responsibility 'no secrets.' No secrets is plausible. It accords with common sense and is not as far reaching as the epistemic partiality of friendship, on which the good friend's beliefs about

their friends involves a "relative ... imperviousness to new evidence" that casts a bad shadow over their friends (Stroud 2006, 514).

What explains no secrets? One consideration is that individuals enjoy a right to privacy whereby evidence about their personal life, including facts about their character, mentality, or personal thoughts (one's deep desires, preferences, etc.) prima facie ought to be disclosed only voluntarily. While transparency about the intimate seems fully optional—there is something unnerving about forming beliefs about intimate matters in cases of involuntarily third-party disclosure for celebrities, even—it seems especially bad in the case of one's friends. This is plausibly due to the intimate nature of the relationship; friendship is constitutively intimate in a way that relationships between acquaintances, colleagues, or strangers are not (or at least need not be), but also the intimate nature of the subject matter: it is information about the person's mental life that she deliberately does not disclose, even to her closest intimates. This suggests that relying on secrets—understood as a certain kind of evidence—to form beliefs about one's friend transgresses a responsibility one bears to their friends, namely, the duty of respect.

It's an open question whether no secrets entails the thesis of epistemic partiality, that friendship (or other intimate relationships) demands partiality in how one evaluates the evidence bearing on the intimates "action or character", namely in a way that favors thinking well of one's intimates (Stroud 2006, 512). The reason why is that epistemic partiality might be thought of as a specific kind of special epistemic responsibility, something that one could reject whilst holding that there are other special epistemic responsibilities. For example, being a good friend might mean that one gives due attention to their friend's fringe arguments, but not necessarily partiality towards the conclusions of their arguments. Epistemic partiality, as Stroud conceives of it at least, looks like a very particular sort of special epistemic responsibility, namely that friends ought to evaluate serious beliefs about each other in ways that are biased favorably towards thinking well of one another.

Torture and Secrets makes the exclusionary rule look plausible: it's permissible to bracket certain kinds of evidence in deliberating about whether *p* because of certain reasons and responsibilities that come out of general respect for persons but also special responsibilities we bear to our intimates. The exclusionary rule helps bring my own view into sharper relief. The exclusionary rule illuminates cases like Influential Philosophy Article if theoretical evidence is within the range of what can be excluded for ethical or broadly prudential reasons. This principle says that:

> *Personal Exclusion Rule*: If P is a non-theoretical personal conviction of *S*, then *S* may bracket theoretical evidence E_T bearing on whether P.

For example, if you believe that "my friends are nice people", but you are unable to identify any flaw in, say, a theoretical argument for other minds skepticism, the personal exclusion rule says that you may bracket that theoretical argument. This helps protect one's personal belief from revisionary

theory. But is also means that one may bracket theoretical evidence which *favors* their personal non-theoretical beliefs as well. So, a theoretical argument for why we have justification for other minds may also be bracketed; one is permitted to rely solely on their prior non-theoretical evidence.

Why does the exclusionary rule hold in cases where theory pressures personal belief? Goldberg (2022) is instructive here. He calls 'contaminated evidence' evidence which "reflects the prevailing prejudices" of one's community, but which one "is not in a position to discern", easily discern, or ought to know reflects prejudices (Goldberg 2022, 387).

For example, consider someone who's news exposure tends to be from sources that over-represent Black men as criminals. Perhaps there are never any stories of Black men in demanding careers, for example. In turn, the viewer will tend to have evidence that biases them against believing a Black man's testimony on more 'intellectual' issues (compare Goldberg 2022, 391). When your total evidence is contaminated, the specific contaminated parts ought to be bracketed. This reflects an "ethical or justice-based constraint" on how we use the evidence we have (Goldberg 2022, 398):

> *Doctrine of Contaminated Evidence*: One's evidence E is contaminated when, by believing P on E, one violates an ethical or justice-based obligation (see Goldberg 2022, 396).

Goldberg focuses on cases in which what one believes is that a certain speaker has evidence about another person's trustworthiness, but because the evidence is contaminated with racial, gender, or other types of social identity prejudices, one thereby wrongs the person by believing that their untrustworthy on the basis of their evidence. When one's body of evidence is contaminated, one ought to bracket the contaminated part. In turn, one should rely on one's total non-contaminated evidence for the target proposition.

The exclusionary rule and the doctrine of contaminated evidence give us a model for understanding why we can exclude theoretical considerations which press on our personal beliefs, but it remains unclear where the ethical permission originates. Is theoretical evidence contaminated? There is an important disanalogy between our cases and the ones that Goldberg is interested in, then. Certain bodies of evidence—those containing certain statistical regularities, internalized stereotypes, or misleading testimony—are contaminated partly in virtue of their relationship to a racist, sexist, ableist, and classist world that forms that evidence. It's a bad feature of social reality. But what's specifically morally problematic is *using* that kind of evidence to form, revise, or maintain beliefs because of the injustices it perpetuates, and the specific harms it causes to certain individuals and groups. Likewise, using theoretical evidence, albeit not necessarily contaminated in the way of reflecting various social ills and injustices, can be used in unethical ways. The doctrine of contaminated evidence doesn't presuppose that the evidence itself contaminated (although it doesn't deny it either). It rather tells us that if believing that p on E would violate an ethical or justice-based obligation, then one should not use the evidence for doxastic deliberation. The source of the

contamination is open-ended. It could be due to the fact that there are various *-ism* prejudices and structural injustices that shape the evidence. Or not. In theoretical cases, it's the latter: there's nothing per se morally objectionable about the evidence, only in how the evidence is prospectively employed. This brings it closer to the legal case. It's not that illegally obtained evidence is itself contaminated but that using it to reach verdicts would be wrong.

So, where do the wrongs originate in the case of using theoretical evidence to form, revise, or maintain personal non-theoretical beliefs, like "I love my partner", "my friends are nice people", "my children's lives are valuable", and so forth? So far, I've appealed to common sense intuitions: ceteris paribus, it just seems wrong—a wrong done to one's partner—to deliberate about whether their partner loves them on the basis of theoretical considerations alone; ceteris paribus, it just seems wrong—a wrong done to one's children—to deliberate about whether their lives have value on the basis of theoretical considerations only.

However, we can also appeal to the effects of making these kinds of deliberations known. While one might think that it's just the revelations that wrong, this is implausible. It's not the telling that does the harm in these cases, but what is antecedent to the telling; it lies in the act of thinking the evidence over that does the wronging (and so it is not like the kind of harm involved in revealing an overheard secret when there is nothing wrong with having overheard the embarrassing secret). After all, if the wrong was due to the revelation alone, then it would be difficult to explain how, if Sarah were to lie ("I ignored an argument against love"), it would not have the moral status of a mere 'white lie'; a kind of petty harm. But the wrong is serious. Although lies wrong to some degree in virtue of their status as lies, some lies are more pernicious than others in virtue what they cover up.

There are other reasons to think that theoretical evidential factoring is wrong in these cases as well. When the belief is tethered to certain commitments, belief revision can lead to dishonoring those commitments, or else to problematic spill-overs. For example, consider the case of an anti-racist activist who begins to suspend judgment about whether racism is wrong. Maybe some theoretical argument she couldn't answer lead her to suspend judgment (bracket for the moment whether she was irrational for this; we're just concentrating on the prospective effects of the attitude change). One effect is that her relationship with her commitment to anti-racism is problematized. Perhaps it will lead her to lapse in both her individual and group anti-racist activities: standing up for a colleague who was discriminated against in their workplace because of their bosses' racial prejudice, or participating in a protest in response to institutional racism. But even if her suspensive attitude doesn't alter these sorts of actions, it can still make her inauthentic in criticizeable ways. The person who protests but doesn't even believe in the cause that their protest embodies is a phony. The person who participates in the kinds of intimate actions characteristic of the good parent, lover,

or friend, but who is agnostic about whether their children's lives matter, they or their partner loves them, or whether their friends are genuine, doesn't align their actions with their thinking; they *practically misrepresent* their mentality. In this way, they misleadingly represent themselves as having the fitting attitudes, speciously rationalizing the positive attitudes—"*Of course* my father thinks my life is valuable" and the like—in their intimates. This seems wrong.

Now, what exactly does 'bracketing' amount to? The concept suggests that the agent is active in surveying their evidence bearing on whether *p*, partitioning their different bodies of evidence, and then judging that *p* on the basis of a refigured mass that excludes the contaminated portion. But this image can't be right, at least not on Goldberg's view, because the agent can be (and often is) ignorant of the contamination, but their ignorance doesn't mean that it's not the case that they ought to bracket the contaminated portion of their total evidence. Let's call this *active bracketing*. Passive bracketing, by contrast, is a matter of not "figuring in" the new evidence; of *possessing* but not *evaluating* the evidence. This is perfectly ordinary. Consider a situation in which one is wondering whether a daily dose of vitamin D is good for their health, and reviews some reliable news outlet *E1* which affirms that it is. In turn, the person becomes confident that a daily dose of vitamin D is good for their health. Later, they skim an article *E2* that says that the research of the effectiveness of a daily dose of vitamin D for health is inconclusive, but they don't evaluate it; it does not enter their system 2 cognitive processes of slower, more engaged self-conscious thinking. Lots of reasons might result in passive exclusion; time-constraints, preferences, background beliefs about what's relevant, heuristics for mitigating risk, and so on.

Of course, decisions might also play a role—"I don't have the time for this", or "I'll carefully review this next week"—but this doesn't mean that the evidential bracketing is now active, since the content of the evidence still remains unevaluated. If Sarah should bracket the theoretical evidence, what this means is that she may not make use of the theoretical evidence as part of her doxastic deliberation. She may continue to base her belief on her prior non-theoretical evidence. She's permitted to keep the theoretical considerations below her serious scrutiny; below the seas of critical reflection.

Before closing, I want to further clarify personalism and its implications. First, one might worry that personalism allows that we can be dismissive towards theoreticians whose arguments we think are mistaken, which would be disrespectful. As McCormick puts it, "[i]f we fail to engage, we fail to treat these believers as rational agents whose acts and attitudes are governed by reasons" (McCormick 2023, 4). While McCormick might be right about this point generally, what personalism says is that one is permitted to not integrate the theoretical evidence into their deliberation—one may pass over it—and that is just one kind of disengagement rather than total withdrawal. It doesn't preclude giving their views attention, discussing their arguments, or listening to their worries; it just

means that one may ignore it for serious, deliberative processes. McCormick's concern is that we don't dismiss people as lost causes, but personalism doesn't lead to such dismissiveness.

Second, one might also worry that personalism threatens the significance of philosophical and humanistic theory if its potential for revision is so severely undercut. I think this is a misunderstanding. One is *permitted* to bracket theoretical evidence when one's personal beliefs are at stake, not *required*. Moreover, unlike the context sensitivity view, it doesn't sever the conceptual relations between highly theoretical and ordinary contexts; personalism simply gives one a series of 'get of jail for free' cards.

Third, one might think that we are just fragmented selves, so that one part of us might believe the revisionary theory while another part of believes the contrary personal belief. It will take us too far afield to adequately assess this move, but it's not so clear what the payoff would be. I don't deny that fragmented belief might be descriptively accurate, only that it is normatively adequate; fragmentation seems like a bad thing. Indeed, personalism is consistent with the fragmentation of belief anyway because it doesn't deny that we couldn't believe both revisionary theories and their personal contraries, only that we may disregard the evidence that would lead us to revisionary theory in the first place.[20]

Fourth, to what extent is personalism inconsistent with evidentialism? Evidentialism is the thesis that, for any rational agent S, the epistemically justified attitude towards *p* at t is whatever their evidence supports at t. In turn, when the revisionary theoretical evidence supports believing some revisionary position, contra a personal conviction, evidentialism thereby allows situations in which revisionary theoretical evidence epistemically justifies revising one's personal conviction.

Still, this possibility doesn't entail that personalism is inconsistent evidentialism, at least understood as only a thesis about epistemic justification and rationality. Personalism doesn't deny that if revisionary theoretical evidence supports believing the revisionary position, contra one's personal conviction, then one would be epistemically justified in adopting the fitting attitudes. It's only inconsistent with the stronger evidentialism that says that what one ought to believe is a function of one's evidence. Personalism severs the link from one's being epistemically justified in taking attitude *A* to *p* to one thereby ought to take *A* to *p*, and in two different ways. One is through bracketing. If one may bracket a body of evidence *E* bearing on whether *p* and heeds that permission, one in effect blocks *E*'s doxastic justificatory power for them with respect to *p*. After all, *E* won't make its way into one's deliberative processing, where one weighs the available evidence; one is, in a morally praiseworthy myside biased way, setting aside that evidence.[21] Another is through

[20] Thanks to Austin Baker for discussing and pressing me on this issue.

[21] One might say that facts about what one ought to believe supervene only on the facts about what the evidence supports, independently of whether the agent possesses and evaluates the evidence. Although there is

moral risk. Consider the stronger evidentialism, and suppose one ought to believe, say, other minds skepticism in response to the skeptic's argument for it. Then one would need to lose their belief that there are other minds and, with it, the beliefs whose truth-conditions depend on it, like (G) "my grandmother cares about me", or (C) "I have children that I love". One ought not believe (G) or (C) if one ought to believe other minds skepticism. But clearly giving up beliefs like (G), (C), and countless others poses serious (even if prima facie) moral risks; it risks altering one's conception of the intimates with whom one interacts; it doesn't exclude representing them as *non-agents*, and so on. That is not a risk one should make; in turn, personalism may be inconsistent with strong evidentialism.

Fifth, one might wonder what personalism means for practitioners in theoretical domains, like philosophy, especially when one is tempted towards extremely revisionary theories. Consider external world skepticism. If I were to believe it, then by own lights I wouldn't be justified in believing that I have children, for that presupposes that there are external things. But then I'm not justified in believing that my son wants my attention—my prospective belief would commit me to things for which, by my view's lights, I don't have any justification to believe. Clearly there are serious moral hazards here.

While there are more specific ways of making *purportedly* revisionary and *purportedly* personal contraries rationally co-tenable, insofar as we allow that there are genuine conflicts, intuitively we should take ownership for the moral risks our attitudes pose. When a philosopher believes a revisionary theory, the personal and moral hazards one absorbs will be a function of both the kind of revisionary theory it is—metaphysical, moral, political—and one's network of interpersonal and practical commitments. In this way, the fitting advice and attendant cautions and criticisms will invariably depend on individual cases.

However, philosophers also have disciplinary responsibilities. Maybe we ought to consider even what might be morally hazardous theories seriously? (Likewise for social theorists, political theorists, etc.) We play a variety of different roles, after all. But we should bear in mind that some of them are more important than others, or at different times. If a new wave of, say, political philosophy takes a turn towards adopting extremist positions, the cost of my doing philosophy might be too great a cost if I'm also politically aspirational. If, given the kind of person I am, I'm attracted to revisionary epistemology and metaphysics, then certain other features of my personal life might ask me to weaken my attention, or turn my research towards other areas.

---

some legitimate sense of ought here—a third-personal and retroactive sense—it is not the strong, normatively guiding sense of 'ought' we're interested in here, where one is pressured to do something.

Personalism helps us sail these waters more safely. How one should exercise their permission will come down to individual cases. While some theories are serious hazards, others may be benign or only mildly risky. Like any hazard, the moral hazard of doing philosophy comes in degrees: is believing external world skepticism closer to crossing a minefield, or more like drinking from a woodland pond? What about pro-cloning, or modal realism? That depends as much on the theory as it does its holder.[22]

Bibliography

Alston, William. (1996). "Belief, Acceptance, and Religious Faith." *Faith, Freedom, and Rationality* (J. Jordan and D. Howard-Snyder (eds.). Rowman and Littlefield: 3–27.

Basu, Rima (2018). The wrongs of racist beliefs. *Philosophical Studies* 176 (9):2497-2515.

Battaly, Heather (2018). Closed-Mindedness and Dogmatism. *Episteme* 15 (3):261-282.

Benatar, David (2017). *The Human Predicament: A Candid Guide to Life's Biggest Questions*. New York: Oxford University Press.

Benatar, David (2006). *Better never to have been: the harm of coming into existence*. New York: Oxford University Press.

Calhoun, Cheshire (1995). Standing for something. *Journal of Philosophy* 92 (5):235–260.

Christensen, David (2013). Epistemic Modesty Defended. In David Christensen & Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press.

Christensen, David (2007). "Epistemology of Disagreement: The Good News". *Philosophical Review*, 116: 187–218.

De Cruz, Helen. (2020). Believing to belong: Addressing the novice-expert problem in polarized scientific communication. *Social Epistemology*, 34(5), 440–452.

Eikenberry, A. M., & Mirabella, R. M. (2018). Extreme philanthropy: Philanthrocapitalism, effective altruism, and the discourse of neoliberalism. *Political Science & Politics*, 51(1), 43-47.

Funkhouser, Eric (2020). A tribal mind: Beliefs that signal group identity or commitment. *Mind & Language*, 1–22. https://doi.org/10.1111/mila.12326

Funkhouser, Eric (2017). Beliefs as signals: A new function for belief. *Philosophical Psychology*, 30(6), 809–831.

Gelpi, R., Cunningham, W., & Buchsbaum, D. (2020). Belief as a non-epistemic adaptive benefit. *Behavioral and Brain Sciences*, 43, E36.

[22] Acknowledgements: I'd like to thank audiences at the European Epistemology Network 2022, the University of Oslo, and the University of California, Irvine (Knowledge, Technology, & Society lectures 2022) for helpful remarks on earlier version of this paper. Thanks also to Mark Walker, René van Woudenberg, Jelle de Boer, and Jakob Olhurst for their helpful feedback.

Gilligan, J. (2007). Terrorism, fundamentalism, and nihilism: analyzing the dilemmas of modernity. *The future of prejudice: psychoanalysis and the prevention of prejudice*, 37–59.

Greaves, H., & MacAskill, W. (2021). The case for strong longtermism. *Global Priorities Institute, University of Oxford*.

Goldberg, Sanford (2022). What is a speaker owed? *Philosophy and Public Affairs* 50 (3):375–407.

Goldberg, Sanford (2013). Defending Philosophy in the Face of Systematic Disagreement. In Diego E. Machuca (ed.), *Disagreement and Skepticism*. Routledge. pp. 277–294.

Goldberg, Sanford (2009). Reliabilism in philosophy. *Philosophical Studies* 142 (1): 105–117.

Heiphetz, Larisa ; Landers, Casey & Van Leeuwen, Neil (2021). Does "Think" Mean the Same Thing as "Believe"? Linguistic Insights Into Religious Cognition. *Psychology of Religion and Spirituality* 13 (3): 287-297.

Jackson, Elizabeth (2021). A Defense of Intrapersonal Belief Permissivism. *Episteme* 18 (2):313–327.

Kahan, Dan M. (2015). Climate-science communication and the measurement problem. *Political Psychology*, 36, 1–43.

Kelly, Thomas (2008). Common sense as evidence: Against revisionary ontology and skepticism. *Midwest Studies in Philosophy* 32 (1):53-78.

Kelly, Thomas (2005). Moorean Facts and Belief Revision, or Can the Skeptic Win? *Philosophical Perspectives* 19 (1):179-209.

Kawall, Jason (2010). Autonomy, agency, and the value of enduring beliefs. *Canadian Journal of Philosophy* 40 (1):pp. 107-129.

Lasonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research* 88: 314–45.

Maguire, Barry and Woods, Jack (2020). The Game of Belief, *Philosophical Review* vol. 129, no. 2.

Marty, M. E., & Appleby, R. S. (Eds.). (1994). *Fundamentalisms observed* (Vol. 1). University of Chicago Press.

Schleifer McCormick, Miriam (2023). Engaging with "Fringe" Beliefs: Why, When, and How. *Episteme,* 1-16.

Moss, Sarah (2018). Moral Encroachment. *Proceedings of the Aristotelian Society* 118 (2):177-205.

Paul, Sarah K. & Morton, Jennifer M. (2018). Grit. *Ethics* 129 (2):175-203.

Pasnau, Robert (2018). Belief in a Fallen World. *Res Philosophica* 95 (3):531-559.

Rinard, Susanna (2013). Why Philosophy Can Overturn Common Sense. In Tamar Szabo Gendler & John Hawthorne (eds.), *Oxford Studies in Epistemology Volume 4*. Oxford University Press. pp. 185.

Stroud, Sarah (2006). Epistemic partiality in Friendship. *Ethics* 116 (3):498-524.

Unger, Peter (1980). Skepticism and Nihilism. *Noûs* 14 (4):517-545.

Unger, Peter (1979). I Do Not Exist. In Graham F. Macdonald (ed.), *Perception and Identity*. Cornell University Press.

Unger, Peter (1971). A Defense of Skepticism. *Philosophical Review* 80 (2):198-219.

Van Leeuwen, Neil & Lombrozo, Tania (2023). The Puzzle of Belief. *Cognitive Science* 47 (2):e13245.

Van Leeuwen, Neil; Weisman, Kara & Luhrmann, Tanya (2021). To Believe is Not to Think: A Cross-Cultural Finding. *Open Mind* 5: 91-99.

Van Leeuwen, N. (2022). Two concepts of belief strength: Epistemic confidence and identity centrality. *Frontiers in Psychology*, 13.

Van Leeuwen, N. (2014). Religious credence is not factual belief. *Cognitio*n, 133(3), 698–715.

Williams, Daniel (2021). Signalling, Commitment, and Strategic absurdities. *Mind & Language*: 1-19.

Wolf, Susan (1982). Moral saints. *Journal of Philosophy* 79 (8):419-439.