

FAITH AND TRADITIONS

Lara Buchak

Princeton University

ABSTRACT: One phenomenon arising in epistemic life is allegiance to, and break from, a tradition. This phenomenon has three central features. First, individuals who adhere to a tradition seem to respond dogmatically to evidence against their tradition. Second, individuals from different traditions appear to see the same evidence differently. And third, conversion from one tradition to another appears to be different in kind from ordinary belief shift. This paper uses recent work on the nature and rationality of faith to show that these features can all emerge from individuals acting rationally—in particular, from individuals rationally having faith in the core assumptions of their traditions. One upshot is that we don't need to employ the idea of incommensurability to explain these features.

1. Introduction¹

Faith has not always been treated as a respectable propositional attitude among philosophers. The reasons for doubt regarding faith's respectability are not far to seek. First, faith has been taken by many to be an intrinsically irrational attitude since the person with faith has an improper attitude toward evidence. Second, faith has often appeared to be a useless attitude since there is no effectual role for it in our cognitive psychology that cannot be played by other more respectable propositional attitudes like belief and desire. According to my account of faith, the *risky-commitment* account, having faith requires stopping one's search for evidence and making a commitment—and maintaining that commitment even in the face of counter-evidence.² I have been at pains to argue that even though faith entails a kind of resistance to counterevidence, this attitude can be rational under certain conditions. In other words, I have defended (and will continue to defend here) the claim that faith is not an intrinsically irrational attitude.

¹ This paper benefited greatly from comments by Ravit Dotan, A.G. Holdier, and Dan Speak; and from discussions at Rutgers University, the University of Manchester, Boston College, Oxford University, Princeton University, and Azusa Pacific University. It was also made possible through the support of a grant from the John Templeton Foundation (the opinions expressed in this publication are those of the author and do not necessarily reflect the views of the John Templeton Foundation).

² See Buchak (2012), Buchak (2014), Buchak (2017), Buchak (2021).

There is more to be said in favor of faith than that it can sometimes be rational. I will show in this paper that faith can help us understand an important phenomenon arising in epistemic life: allegiance to, and break from, a tradition. This phenomenon has three central features. First, individuals who adhere to a tradition seem to respond dogmatically to evidence against their tradition. Second, individuals from different traditions appear to see the same evidence differently. And third, conversion from one tradition to another is qualitatively different from ordinary belief change: it is discontinuous and felt as a ‘break’ rather than smooth and gradual. I will show that, contrary to initial appearances, these features can all emerge from individuals acting rationally—in particular, from individuals rationally having faith in the core assumptions of their traditions.

2. Traditions and Paradigms

The phenomenon of *paradigms* and *paradigm shift* was first identified by philosophers of science observing that science sometimes undergoes large-scale changes instead of progressing incrementally.³ The rough pattern begins with a scientific community centered around some core assumptions or practices. For example, the scientific community of the 2nd century that accepted Ptolmeic astronomy centered itself around the assumption that the Earth is a sphere at the center of the universe and postulated that each planet was moved by a system of spheres.⁴ Once the core assumptions are accepted by the community, there is little interest in justifying them further; they are instead taken for granted while the community carries out the ‘normal’ scientific work of discovering facts about the world. On some occasion, however, the community is confronted by crisis, in the form of a problem or problems it cannot solve: some anomaly arises that is clearly not predicted by the theory. For example, observations of the positions of the planets seemed to contradict the geocentric view. When a crisis occurs, there are a number of responses by scientists within the community. Some may dig their heels in and make the theory more complex to account for the anomaly, in a way that can seem recalcitrant or dogmatic; for example, geocentrists postulated more and more complex versions of the Ptolmeic theory. Others might search about for new theories, dropping the assumptions of the old paradigm and working out new ones. If the anomaly is not accounted for, there can eventually be a whole-scale shift to an entirely new set of core assumptions. In what is typically known as the ‘Copernican revolution’, the geocentric theory was replaced by theories that held that Earth and the other planets move around the sun. It is significant that

³ Kuhn (1970), Lakatos (1970), Feyerabend (1975).

⁴ See Kuhn (1962).

the term ‘revolution’ is used: a shift to a new set of core assumptions appears qualitatively different than ‘normal’ science.

Following the discussion of this phenomenon in science, philosophers have noticed it in other domains.⁵ Alastair MacIntyre argues that moral traditions similarly center themselves around certain core assumptions, and communities go through a process like the one described above.⁶ For example, Aristotelians centered themselves around a particular conception of the virtues and how to elucidate them.⁷ Within a moral community, there is often little interest in justifying the core assumptions; but when a crisis arises in the form of an encounter with a different tradition, this becomes pressing. As in the scientific case, there may be a variety of responses: for example, discovering new concepts to resolve the inadequacies in the current tradition, rejecting the new tradition as a source of evidence, or shifting allegiance to the new tradition. Nancey Murphy argues that theology also holds core assumptions that are accepted without additional justification, and ‘normal’ theology is carried out in light of these, until there is a crisis. For example, the Catholic modernists centered themselves around the assumption that Catholicism is true and can be reconciled with modern thought, and looked to devotional religious life to find empirical support for theology.⁸

This pattern can also occur not just within a community but within an individual.⁹ For example, consider a person who grew up comfortably as a Christian for whom being an atheist was not a serious intellectual option—or a person who grew up comfortably as an atheist for whom being a Christian was not a serious intellectual option. They each may start with their core assumption or assumptions (that constitute what it is to be a Christian or an atheist), and explore the world using principles that emanate from those assumptions. The Christian might encounter unnecessary suffering, the atheist a lack of meaning—they will both initially be inclined to dismiss these challenges, in a way that looks like dogmatism to observers. Or when they meet each other, they may initially convince themselves that the other is unreasonable,

⁵ See, e.g., Murphy (1990), MacIntyre (1988), Taylor (1985), McDowell (1998), and Smith (2011).

⁶ MacIntyre (1988).

⁷ MacIntyre (1988: 349).

⁸ Murphy (1990: 92, 101). Kuhn himself also notes similarities between science and theology.

⁹ MacIntyre (1977), van Fraassen (2002), Faulkner (2019). Empirical discussions of religion conversion include Lofland and Stark (1965), Gordon (1967), Kox et al (1991), Rambo (1993), and Iyadurai (2010), each of whom describe a tension, dissatisfaction, or dilemma similar to the ‘crisis’ in science. For example, Lofland and Stark (862) note that in order to convert, “a person must experience, within a religious problem-solving perspective, enduring, acutely-felt tensions.”

morally unserious, or intellectually inferior. This may resolve into a renewed adherence to their own traditions. But instead, one or both of them may feel an increasing mismatch between their theory and lived experience, and when the build-up is too much, they might leave their tradition and convert to a different one. It is significant that this is described as *conversion* and not merely a shift in beliefs: it is often felt as a significant break, rather than part of an ordinary, continuous process.¹⁰

Other examples abound, in the domains of morality and politics: a committed consequentialist or deontologist; a pro-choicer or pro-lifer; a republican or democrat. All see the world through a lens which they no longer feel the need to spend time justifying, but sometimes confront crises, in the form of a thoughtful peer who disagrees or a fact they just can't seem to fit into their point of view. Sometimes they dig in, sometimes they experience a conversion to the alternative. There are interpersonal examples as well. You may start off by viewing a friend or a spouse as trustworthy, kind, good, but then confront evidence that they are not. You may initially ignore it completely, then later feel a build-up of evidence that you cannot ignore. Some who experience divorce or a break-up describe it as a seismic shift from one way of seeing their partner to a wholly different way.

I will use the more general *adherence to a tradition* and *break from a tradition* to describe cases that fit the general pattern discussed. The pattern has three features, each of which seem irrational, or at the very least, *arational*. First, adherents to a tradition initially respond seemingly dogmatically—by ‘digging in’—to anomalies or counter-evidence. The geocentrist makes his theory increasingly complex; the Christian creates theodicies; neither becomes less committed to their tradition. Call this feature *recalcitrance*. Second, adherents to different traditions can look at what is putatively the same evidence and draw very different conclusions—or share all their evidence and still draw different conclusions. The pro-lifer and the pro-choicer look at the same world; for one it is filled with babies being killed, for another with women being denied their rights. Call this feature *gestalt*. Third, a break from one's tradition is discontinuous, a drastic response to built-up tension rather than an ordinary change in belief. The converting atheist doesn't gradually accept some Christian claims alongside her atheistic ones; she instead becomes a Christian. Call this feature *conversion*.

¹⁰ This isn't to say that conversion is *temporally* instantaneous rather than temporally gradual; see, e.g., Kilbourne and Richardson (1989) for empirical discussion of both types of conversion. The key point is that conversion, unlike ordinary belief change, resolves tension that is *built up* and involves a drastic epistemic reorientation.

A common way to explain these features is to say that traditions are *incommensurable*: in particular, the concepts used in one tradition differ from those used in another tradition, such that the two traditions ‘don’t make sense’ from each other’s perspective.¹¹ Simplifying a bit, this fact is supposed to explain the three features as follows. Since a tradition includes its own distinctive concepts and vocabulary, one might not even be able to ‘see’ the counter-evidence, so recalcitrance results. The Christian cannot see unnecessary suffering, she can only see the working out of God’s plan. The devoted wife cannot see her husband’s selfish actions as selfish, she can only see them as those of a kind person under stress at work. Furthermore, there is no such thing as un-interpreted evidence, so one’s evidence depends on one’s paradigm: since practitioners of different paradigms can’t actually look at the ‘same’ evidence, gestalt results. Glancing at an early sonogram, the pro-lifer sees a person, the pro-choicer sees a clump of cells, and neither can understand each other. Finally, a shift from one set of concepts and way of seeing to another is different in kind from a shift in beliefs concerning a given set of concepts—hence conversion.

Some have challenged the idea that traditions really can be incommensurable.¹² Furthermore, regardless of whether incommensurability is *possible*, there seem to be cases that can aptly be called cases of paradigm shift or conversion in which incommensurability is not present—both theories are perfectly understandable from the point of view of the other. Practitioners of different moral theories take themselves to understand the alternatives, for example. On the most natural interpretation of what’s going on, they understand the vocabulary, they just disagree with the claims. The pro-lifer and pro-choicer both understand what it is to be a person or merely a clump of cells—they just disagree about whether the fetus is a person or merely a clump of cells. Atheists can have as sophisticated a concept of God as Christians do, they just disagree about whether something instantiates that concept. Furthermore, on at least Kuhn’s view of incommensurability, the new paradigm is supposed to contain language in which we can understand the old one; so if traditions were incommensurable, then you would expect paradigm shift to be uni-directional. But it’s not: atheists become Christians and Christians become atheists, conservatives become liberals and liberals become conservatives, consequentialists become virtue ethicists and virtue ethicists become consequentialists, pro-lifers become pro-choicers and pro-choicers become pro-lifers.

¹¹ This idea originated with Kuhn (1962) and Feyerabend (1962), who focused on scientific paradigms in particular; but it has been imported to ‘individual’ epistemology.

¹² See, for example, Davidson (1973).

Perhaps there are responses to these worries. I am not interested in showing definitively that incommensurability cannot do the work it needs to do. Instead, I want to show that we can explain the familiar pattern, and explain it as *rational*, without invoking incommensurability at all. I will focus on the case of an *individual's* adherence to (and break from) a tradition, but everything I say can be applied to the communal case. I will use my account of faith—the *risky-commitment account*—to show that when an individual rationally has faith in the core assumption of a tradition, all three features follow; and I will argue that core assumptions of paradigms are precisely the kinds of things that are appropriate objects of rational faith.

3. The Risky-Commitment Account of Faith

I here briefly rehearse the risky-commitment account of faith,¹³ and of when and why such faith is rational. This account is an account of *propositional* faith, and although it encompasses both sacred and secular cases of faith, only certain propositions are candidates for faith at all. In order for a proposition to be a potential object of faith for a subject, the subject must care whether the proposition is true, must have a positive attitude towards the proposition, and must not be certain of the proposition on the basis of his evidence alone—his evidence must leave it open that the proposition is false.

Faith is tied to action, and having faith is a matter of being disposed to do certain actions on the basis of the proposition we have faith in: in particular, being willing to take risks on the basis of that proposition. When we have faith that a particular individual will act in a certain way—keep our secret, pick us up from the airport, do what is in our best interests—we take a risk that the individual will let us down. We are vulnerable to the individual not acting as we have faith that she will act, in the sense that we will bear some cost if she lets us down.¹⁴ But not every case of risk-taking is an act of faith. Faith requires a willingness to act on the proposition one has faith in, without first looking for further evidence—and a willingness to continue to act, even if counter-evidence does come in. If one has faith that one's friend will pick one up at the airport, one won't call to remind her, nor will one call a cab if one can't find her. If one has faith that one will complete graduate school, one will continue in the program even after failing an exam. In other words, faith requires making a commitment to take a risk and sticking with that

¹³ See Buchak (2012, 2017, 2021).

¹⁴ McKaughan (2017) also notes that faith involves an active commitment. He uses the term “reliant relational response,” although the account here is meant also to apply in cases in which there is no relevant person or no relevant relationship.

commitment through periods of epistemic doubt. Which of the two aspects of faith—making a commitment or sticking with it—is more central will likely depend on the case.

Thus, a subject has faith in some candidate proposition if he is willing to commit to taking risks on the proposition without examining additional evidence, and to maintain that commitment even in the face of counter-evidence. More formally:

A proposition ‘X’ is a *candidate for faith* for a person S if S has a positive attitude towards the claim that X and is not certain whether X holds on the basis of his evidence alone.¹⁵

S *has faith* that X if and only if:

- (i) ‘X’ is a candidate for faith for S.
- (ii) S is willing to commit to take (subjective) risks on the claim that X, independent of additional evidence and
- (iii) S is willing to follow through on such risky actions even when he receives evidence against the claim that X.¹⁶

To the extent that S is willing to perform riskier acts that express his faith, and remain steadfast in the face of stronger counter-evidence, he has a higher degree of faith that X.¹⁷

Steadfast *believing* is a special case of faith in the proposition believed, provided one has a positive attitude towards the proposition and is not certain on the basis of the evidence alone. Belief is a risk on the truth of the proposition believed, and so if one is willing to commit to believe without further evidence and maintain that commitment even when counterevidence arises, one performs an act of faith by believing in this way. When a person adopts a belief that he is committed to even if counterevidence should arise, I will call that belief a *conviction*. Provided one has a positive attitude towards X, and is not certain that X on the basis of the evidence alone, adopting and maintaining a conviction that X is an act of

¹⁵ There are different ways to fill out what positive attitude is required: see, for example, Howard-Snyder (2013), who requires both a positive evaluation of X and a positive conative orientation towards X, but holds that many different attitudes could stand in for each of these requirements.

¹⁶ Condition (iii) applies after the initial choice but before the risky act has been performed or completed.

¹⁷ See Pace (2017) on degrees of faith.

faith that X. Thus, while faith needn't rest on belief, it can *manifest* as a belief—a belief that is resilient in the face of counterevidence.

Both *acting* in the face of counter-evidence and the more specific *believing* in the face of counter-evidence will be important to our discussion of traditions. Although in some central cases like science, belief is central, in other cases like religion and morality, adhering to a tradition importantly involves not just believing when counter-evidence comes along, but acting in accordance with the dictates of the tradition.

An initial thought is that faith is always irrational, because acting and believing on the basis of more evidence rather than less—and changing one's plans and beliefs as new evidence comes in—better conduces to getting what one wants. However, as I've argued, there are situations in which, because of the structure of the available acts and the available evidence, it is rational to commit to a risky act rather than to examine additional evidence; and it is rational to maintain this commitment even in the face of counter-evidence.¹⁸ These include situations in which the risky act is belief. I turn now to these situations.

4. Rational Faith¹⁹

Formal epistemology concerns itself with two important attitudes: credences and utilities. I will understand credences—an individual's subjective degrees of belief—to express the relationship, according to her, between her evidence and the truth of various propositions: roughly, they express how likely she thinks various propositions are to be true, given the evidence she has. Credences obey the probability calculus, and are updated by conditionalizing on new evidence. The argument here is neutral between a purely subjective view, according to which one can have any internally coherent credence assignment, and a view according to which one must meet additional credence-to-evidence standards in order to count as rational. I will understand utilities—an individual's desires—to be a measure of how valuable consequences are for the individual. Again, I will be neutral between a “Humean” view of

¹⁸ See Buchak (2010) and Buchak (2012) for the argument that it is instrumentally rational to commit rather than to look for more evidence, and see Buchak (2017) for the argument that it is instrumentally rational to maintain the commitment in the face of counter-evidence. See Buchak (2021) for these arguments applied to belief specifically.

¹⁹ Some of the material and results in this section are taken from Buchak (2021).

utility, according to which one can have any internally coherent utility assignment, and a view according to which one must meet additional standards.

To analyze the rationality of holding an ‘on-off’ belief, we can assign *epistemic utility* values to holding that belief in various states of the world. Following William James and adapting a model from Isaac Levi, the utility value of holding a belief will be a function both of that belief’s truth and its informativeness.²⁰ For example, if it is snowing, it is better to believe ‘it is snowing’ than to believe ‘it is either snowing or raining’ (while both are true, the former is more informative), and it is better to believe either of those things than to believe ‘it is raining’ (which is false). In the case of a single proposition X , if X is true then $u(\text{believe ‘}X\text{’}) > u(\text{believe ‘}X \vee \text{not-}X\text{’}) > u(\text{believe ‘not-}X\text{’})$. We can, without loss of generality, assign a value of 1 to believing an informative truth, a value of 0 to believing a falsehood, and some middle value M (with $0.5 \leq M < 1$) to remaining agnostic, that is, to believing the non-informative truth ‘ X or not- X ’:²¹

	X	Not-X
Believe ‘ X ’	1	0
Be agnostic	M	M
Believe ‘not- X ’	0	1

FIGURE 1. Payoffs of Various Belief States, $0.5 \leq M < 1$

²⁰ Here, ‘informativeness’ refers to information *content*, i.e., how much the belief pares down the possibilities or how ‘specific’ the belief is, rather than information *value*, i.e., how much the agent values information about the particular topic. James (1896), Levi (1967a), Levi (1967b). In Levi (1980) and Levi (1991), Levi embeds this decision-theoretic framework in a complex view about knowledge; in this paper I primarily make use of the decision-theoretic framework, which can be embedded in a variety of views about knowledge. Related contemporary views that postulate that rational belief maximizes epistemic utility include Easwaran (2016) and Dorst (2019).

²¹ The utility function thus takes as its input the ‘cognitive’ or ‘epistemic’ consequences of believing a particular proposition in a particular state of the world, and outputs a real number in $[0, 1]$. When we introduce additional values b and d in the subsequent discussion, the range of values that the utility function can take on will widen to $[-d, 1 + b]$.

According to the standard view (expected utility maximization), an individual is instrumentally rational if she chooses actions that maximize expected utility, according to her own credences and utilities. So, if one has credence $p(X) > M$, then one ought to believe 'X'.²²

What picture of belief and its epistemic function am I adopting here? I want to remain neutral between several options.

The first option is a picture on which an epistemic agent assigns credence to all propositions, and then additionally adopts an on-off epistemic attitude—belief—towards some of them, on the basis of these credence assignments and in accordance with the epistemic utility framework detailed above. In typical versions of this picture, one does not assign credence 1 to a proposition unless one considers one's credence unrevisable.

This is not Levi's picture, however. For Levi, an epistemic agent first adopts a corpus of propositions that she accepts 'as evidence,' i.e., that are not subject to serious question, though are not immune from revision in the future; these propositions constitute her *knowledge*.²³ After doing so, she assigns credences to the remaining propositions on the basis of this corpus; she then rationally accepts propositions that maximize epistemic utility as detailed above and adds these to her corpus; and she iterates this process.²⁴ At each step, credence essentially 'drops out', since once a proposition is accepted into her corpus, new credences are calculated.

The second option is to accept a picture like Levi's, but without iteration and, more importantly, without throwing out the credences that were based on the initial corpus. On this picture, like Levi's, there are some propositions that the agent adopts in a 'strong' way—they constitute knowledge or evidence or their negations are not serious possibilities—and these are used to assign credences to the remaining

²² In the case of ties, Levi holds that we should privilege agnosticism over belief, and I will follow this rule (Levi 1967a: 84). Note that Levi's view is not the 'threshold' view (according to which belief is credence above a threshold), because when multiple propositions are at issue, the minimum credence required to believe 'X' will depend on both the propositions involved and on what else the subject accepts.

²³ Roughly, Levi favors the view that the corpus described above (an agent's 'knowledge') serves as her standard for serious possibility; she thus will not acknowledge the possibility that items in her corpus are possibly false; however, knowledge is *corrigible* in that an agent may come to regard something that is now in her corpus as possibly false later (Levi 1983: 13-19).

²⁴ See, e.g., Levi (1983: 51-54).

propositions.²⁵ But *in addition*, the agent accepts some of these remaining propositions in a ‘weak’ way—she ‘merely’ believes them, but does not know them or take herself to know them or treat them as evidence—in accordance with the epistemic utility framework.

Each picture holds that belief is an on-off attitude that outruns what one assigns credence 1 to or what one knows or takes as evidence. This attitude allows us to represent the world as including the truth of some claim, even if we do not know or do not assign credence 1 to that claim. Proposals in the literature for why we need this attitude include: that this attitude forms the basis for reactive attitudes like praise and blame; that it forms the basis of assertions; that it is action-guiding; and that it is reason-guiding.²⁶ These proposals generally hold that while decision theory guides many parts of practical and epistemic life, other parts of practical and epistemic life must be guided by norms that use beliefs rather than credences. Adopting a belief doesn’t change the underlying credence calculations: one doesn’t, for example, assign probability 1 to things that one believes. Instead, beliefs co-exist with credences and play a different role in epistemic and practical life.

Having beliefs, on these pictures, is a matter of representing the world in a certain way. Furthermore, beliefs are open to revision at any given time. A key point for my argument in this paper, which I will turn to shortly, is that it is sometime beneficial to *commit* to a belief—to have a *conviction*. (Adopting the second picture above, we could say: it may be beneficial to adopt an attitude that plays the same role as knowledge or evidence or serious possibility—in that it is stable and not normally a candidate for current revision—but when one’s evidence doesn’t warrant the stronger attitude.)

So, we have: credences (perhaps against the background of a body of strongly-accepted propositions), which represent the degree to which evidence supports various propositions; beliefs, which represent the propositions that the agent takes to be true, even though she may be uncertain of them or take their negations to be serious possibilities; and convictions, which represent the beliefs she is committed to holding even in the face of counterevidence. We will ask: if there are benefits to *true convictions*—not just to true beliefs—then how does this affect what we should believe over time, and how our beliefs should relate to the evidence?

²⁵ This picture so far is roughly the picture found in Williamson (2000: ch. 10), in which one’s evidence is the propositions one knows, and then probabilities are assigned on the basis of these propositions.

²⁶ See Jackson (2020) for a helpful overview of this literature.

I will first argue that there are in fact benefits to having a true conviction over and above the benefits of having a true belief at a time. I will then use the formal model to show that, surprisingly, when there are such benefits, it can be rational to let your beliefs ‘swing free’ of the evidence at certain points in time. Thus, I will show that the value of committing to *true* beliefs can make it rational to commit to beliefs whose truth is in doubt.

There are at least three ways in which it can be beneficial to be committed to a belief, if that belief turns out to be true. First, there might be matters on which believing the truth at every point in time is better than sometimes believing the truth and sometimes suspending judgment or believing a falsehood, over and above the value of a true belief at each time—it might be better to have *unwavering belief* in the truth. Interpersonal matters are often like this: it is better, for example, to unwaveringly believe of a trustworthy spouse that she is trustworthy, rather than to sometimes believe this and sometimes not. Religious or moral views often involve an interpersonal element as well: for example, it is better to unwaveringly believe of God that he loves you rather than to sometimes believe he is indifferent.

A second benefit of committing to a belief over time (if that belief is true), over and above the benefit of merely holding that belief at a given time, is what we might call the benefit of *confident evaluation*.²⁷ When we are committed to a belief, we reason as if it is true. We cut down on the number of cognitive options that are ‘live’ at any given moment. We ignore sources that reason from claims incompatible with the belief, freeing up our intellectual resources to use in other ways. We pay special attention to sources that reason from the belief, to better absorb existing evidence. For example, a committed virtue theorist ceases to pay attention to consequentialism or deontology: she will spend her time reading and thinking about the virtues and how to attain them, rather than about how to do the most good possible or keep the commands. A committed democrat ceases to pay attention to libertarian sources: he will be interested in the debate surrounding which policies best care for the marginalized, rather than that surrounding which policies are most consistent with liberty. We collaborate with others who share our belief, to discover new evidence. Two committed Christians might pray together and report on what they think God is telling them. We come up with larger theories on the basis of the belief. Someone committed to the idea that a fetus is a person will arrive at moral principles about what we owe expectant

²⁷ See also Holton (2004, 2014). Here we possibly depart from Levi’s model, since his is more holistic in nature, e.g., evaluating all of the atomic propositions one cares about.

mothers and which social arrangements and laws are just. It is useful to commit to a claim that helps us reason in these ways, and it is expensive to drop this commitment, since one then has to start anew on all of these tasks.²⁸ (Thus, these benefits can often only be realized by making a commitment that is costly to drop later.)

Finally, committedly holding a true belief across time may be better than merely holding a true belief at each time when one is engaged in a project that requires a series of actions over time in order to be successful, insofar as belief facilitates action. Moral and religious projects both seem to have this character: working towards justice, working to get money to the most effective charities, living as a nun, engaging in the normal practices and rituals of a religious tradition. These are all projects that require action over time, and that action is much easier to perform—especially in the face of temptation—when one believes the claims on which one’s work is based and when one can rely on believing them in the future.

Although this paper is primarily about belief, this last point also applies to the benefits of committed *action* rather than mere action-at-a-time. It is sometimes better not just to do *A*, but to be committed to doing *A*. Again, one might be engaged in a project that requires a series of actions over time, and it is better both for purposes of planning and for purposes of carrying out the project (if it will in fact succeed) that one commit to performing the relevant actions at each time. Alternatively, someone else’s actions might be based on your commitment to doing *A*: for example, I can more easily make decisions on the basis of my husband’s *commitment* to stay in our marriage than on the basis of his current, but revisable, plan to stay in our marriage.

If I am correct that there can be benefits to committing to a belief if that belief is in fact true, then we can evaluate when rationality requires us to commit to believe ‘X’ regardless of future evidence (condition (ii) of having faith that X) and when rationality requires us to in fact follow through and remain committed when we in fact receive evidence against X (condition (iii) of having faith that X).

²⁸ Some have focused on the idea that by dividing cognitive labor among individuals, the group can benefit, and so it is helpful for individuals to be committed to different views, even though some will be committed to false views (Kitcher (1990), Strevens (2003), Elgin (2010), Muldoon (2013), Fleisher (2018), Loughheed (2018), Kraay (2019)). But here I note the benefits to the *individual*, if she is committed to a *true* belief.

Assume that one is currently rationally required to believe ‘X’, that is, that $p(X) > M$. And assume that the evidence one will receive is either E (‘positive’ evidence that raises one’s credence in X) or not-E (‘counter-evidence’ or ‘negative’ evidence that lowers one’s credence in X). Assume further that if one receives not-E, then one’s credence in ‘X’ will be low enough that one will be rationally required not to believe ‘X’: $p(X | \text{not-E}) \leq M$. Assume finally that if one receives not-E, one will be rationally required to remain agnostic rather than to believe ‘not-X’: $p(\text{not-X} | \text{not-E}) \leq M$. Taking these assumptions together: $1 - M \leq p(X | \text{not-E}) \leq M$.

With these assumptions, the act of basing one’s belief on further evidence has the effect of yielding utility M if one receives negative evidence, since negative evidence will make one adopt agnosticism: one will get M regardless of the truth of ‘X’. On the other hand, the act of committing to believe ‘X’ has the effect of yielding utility 1 if one receives negative evidence but X holds nonetheless (if the evidence is ‘misleading’) and yielding utility 0 if one receives negative evidence and not-X in fact holds. Furthermore, if there are benefits to commitment, then the act of committing to believe ‘X’ yields an additional benefit in all of the states in which X holds. We can represent this in matrix form:

	X & E Correctly leading positive evidence	X & not-E Misleading negative evidence	Not-X & E Misleading positive evidence	Not-X & not-E Correctly leading negative evidence
Commit to believe ‘X’	1 + b	1 + b	0	0
Believe ‘X’ and re-evaluate in response to negative evidence	1	M	0	M

FIGURE 2. Payoffs of Committing to Believe ‘X’ versus Re-evaluating, with Commitment Benefits

We can establish the conditions under which it is better (has a higher expected utility) to *commit* to believe ‘X’ rather than to believe ‘X’ but plan to re-evaluate if one receives negative evidence.²⁹ The

²⁹ Committing will be better than believing but planning to re-evaluate if the following value is positive: $EU(\text{commit}) - EU(\text{believe \& re-evaluate}) = p(\text{not-E})[p(X | \text{not-E}) - M] + p(X)b$, i.e., as long as $p(X)b + p(\text{not-E})(p(X | \text{not-E})) > p(\text{not-E})M$. (Keeping in mind that $p(X | \text{not-E})$ must be less than or equal to M but greater or equal to $1 - M$, so that one will prefer agnosticism upon receiving counter-evidence; if it is too high one will prefer to believe ‘X’ and if it is too low one will prefer to believe ‘not-X’). One limit case in which this is positive is when $p(X | \text{not-E}) = M$ and b is positive ($p(X)$ is anything that satisfies the initial requirement that $p(X) > M$). As $p(\text{not-E})$ increases and $p(X | \text{not-E})$ decreases, we require a larger $p(X)$ and/or a larger b for $EU(\text{commit}) - EU(\text{believe \& re-evaluate})$ to remain positive.

basic idea is that commitment has an inherent benefit if X holds, and commitment additionally keeps one from being swayed by misleading negative evidence—but it also (unfortunately) keeps one from being swayed by correctly leading negative evidence. So insofar as the inherent benefits of commitment (if X in fact holds) are high, or the probability of misleading negative evidence is not small compared to the probability of correctly leading negative evidence—being misled into dropping one’s true belief is a significant possibility—committing trumps a plan to re-evaluate.

If one in fact receives counter-evidence not- E , one faces a new decision, namely whether to stick with the commitment or back out. We can think of backing out in one of two ways, or a combination thereof. First, we can think of backing out as merely forgoing the benefits of commitment: one would have gotten $1 + b$ if one had stuck with the commitment to believe ‘ X ’, but instead one gets the normal payoff of agnosticism, M . Second, we can think of backing out as paying a penalty: instead of receiving the normal payoff of agnosticism, one receives a lower payoff, $M - d$. (A penalty for backing out might arise, for example, from the fact that one has to re-learn information one discarded or ignored when one was committed to ‘ X ’; for example, the committed democrat who backs out might have to re-learn which policies are best if libertarianism is true.). Here is the choice, with both options represented.³⁰

	X & not-E Misleading negative evidence	Not-X & not-E Correctly leading negative evidence
Believe ‘X’ (stick with commitment)	$1 + b$	0
Be agnostic (abandon commitment)	$M - d$	$M - d$

FIGURE 3. Payoffs after Negative Evidence, with Commitment Benefits

³⁰ Setting $d = 0$ represents that there is no penalty for backing out of the commitment, and thus can represent the thought that abandoning the commitment merely involves forgoing the benefits of the commitment; setting $b = 0$ represents that there are no (further) benefits to sticking with the commitment—the benefit has already been ‘consumed’—and thus can represent the thought that abandoning the commitment merely involves paying a penalty for backing out. Assigning positive values to both b and d represents the thought that one both forgoes some benefits, and pays some costs, when abandoning the commitment.

Thus, we can also establish the conditions under which it is better (has a higher expected utility), once one has received counter-evidence, to remain committed to believing ‘X’ rather than to in fact re-evaluate.³¹

I have just discussed the conditions under which commitment can be rational, under the assumption that the new evidence might lead one to be agnostic in the absence of a commitment. But one might also wonder whether and under what conditions it can be rational to commit to believe ‘X’ if the new evidence might lead one to believe ‘not-X’—rather than simply be agnostic—in the absence of a commitment. As it turns out, this can be rational, and the conditions are similar to those for the case in which new evidence might lead one to be agnostic.

Let us assume that negative evidence will be strong enough to make it rational to believe ‘not-X’: $p(\text{not-X} \mid \text{not-E}) > M$, i.e., $p(X \mid \text{not-E}) < 1 - M$. Then the choice between committing to believe ‘X’ and basing one’s belief on further evidence has the effect of yielding the payoffs for ‘not-X’ if one receives negative evidence, and the payoffs for ‘X’ (without commitment) if one receives positive evidence:

	X & E Correctly leading positive evidence	X & not-E Misleading negative evidence	Not-X & E Misleading positive evidence	Not-X & not-E Correctly leading negative evidence
Commit to believe ‘X’	1 + b	1 + b	0	0
Believe ‘X’ and re-evaluate in response to negative evidence	1	0	0	1

FIGURE 4. Payoffs of Committing to Believe ‘X’ versus Re-evaluating, with Commitment Benefits and Strong Negative Evidence

Again, we can establish the conditions under which it is better (has a higher expected utility) to commit to believe ‘X’ rather than to plan to re-evaluate in response to new evidence.³² And again, it will be better to

³¹ It will be better to remain committed as long as $p(X \mid \text{not-E}) > \frac{M-d}{1+b}$. See Buchak (2021). Intuitively: if the benefits of commitment are higher, or the costs of backing out higher, then commitment will remain the best option for lower values of $p(X \mid \text{not-E})$.

³² Committing will be better than believing but planning to re-evaluate if the following value is positive:
 $EU(\text{commit}) - EU(\text{believe \& re-evaluate}) = p(\text{not-E})[2p(X \mid \text{not-E}) - 1] + p(X)b$, i.e., when $p(X)b + 2(p(\text{not-E})p(X \mid \text{not-E})) > p(\text{not-E})$. (Keeping in mind that $p(X \mid \text{not-E}) < 1 - M$, so that one will prefer to believe ‘not-X’ upon

commit insofar as the inherent benefits of commitment (when X holds) are high, or the probability of misleading negative evidence is not small compared to the probability of correctly leading negative evidence (i.e., being misled into dropping one's true belief is a significant possibility).

We can similarly consider the new decision one faces if one in fact receives strong counterevidence: the decision of whether to stick with the commitment or back out and instead believe 'not-X'. As in Figure 3, we can represent abandoning the commitment either as forgoing the benefits of the prior commitment or as paying a penalty:

	X & not-E Misleading negative evidence	Not-X & not-E Correctly leading negative evidence
Believe 'X' (stick with commitment)	1 + b	0
Believe 'not-X' (abandon commitment)	0 - d	1 - d

FIGURE 5. Payoffs after Negative Evidence, with Commitment Benefits and Strong Negative Evidence

And again, we can establish the conditions under which it is better, once one has received counter-evidence, to remain committed to believing 'X' rather than to switch to believing 'not-X.'³³

To summarize: in both cases—the case in which counter-evidence would make one agnostic in the absence of the commitment and the case in which counter-evidence would make one believe 'not-X' in the absence of the commitment—there are two conditions that must be met in order for commitment to be rational. The first is a condition that is met to the extent that the benefits of commitment are high, the probability of X is high, and the probability of misleading evidence is high:

$$\text{Agnosticism case: } p(X)b + p(\text{not-E})(p(X | \text{not-E})) > p(\text{not-E})M$$

$$\text{Belief in 'not-X' case: } p(X)b + 2(p(\text{not-E})p(X | \text{not-E})) > p(\text{not-E})$$

receiving counter-evidence.) Again, as $p(\text{not-E})$ increases and $p(X | \text{not-E})$ decreases, we require a larger $p(X)$ and/or a larger b for $EU(\text{commit}) - EU(\text{believe \& re-evaluate})$ to remain positive.

³³ It will be better to remain committed as long as $p(X | \text{not-E}) > \frac{1-d}{2+b}$. One could add an option in which one commits to 'not-X' and thus receives an additional benefit if not-X holds; this option's relative value would again depend on the evidence one might receive in the future, so full discussion of this possibility is beyond the scope of this paper.

The second is a condition that is met to the extent that the benefits of commitment are high, the costs of backing out are high, and probability of misleading evidence is high:

$$\text{Agnosticism case: } p(X | \text{not-E}) > \frac{M-d}{1+b}.$$

$$\text{Belief in 'not-X' case: } p(X | \text{not-E}) > \frac{1-d}{2+b}.$$

Putting these two conditions together, we can say roughly that one is rationally required to commit to believe 'X', and to follow through on that commitment when one gets negative evidence, even though one wouldn't have high enough credence to believe 'X' after getting the negative evidence if one hadn't made the commitment, when:

- (i) $p(X)$ is sufficiently high: one already has significant evidence for the belief in question.
- (ii) $p(X | \text{not-E})$ is sufficiently high: being misled into dropping one's true belief is a significant possibility.
- (iii) There are benefits to be gained from committing to believe 'X', if X holds, that couldn't be gained from merely believing 'X', and one must actually follow through on the commitment in order to get these benefits (or to avoid paying the costs of backing out).

What counts as 'sufficiently high' depends on the value of commitment and the price of backing out, and (in the case in which one might switch to agnosticism rather than to believing 'not-X') the value of agnosticism relative to belief in the truth and belief in a falsehood.³⁴

³⁴ Here are some examples, to give the reader a sense of the magnitude of the values:

Cases of rational commitment, in which counter-evidence would lead to agnosticism in absence of a commitment (taken from Buchak (2021): $M = 0.75$, $p(X \& E) = 0.85$, $p(X \& \text{not-E}) = p(\text{not-X} \& E) = p(\text{not-X} \& \text{not-E}) = 0.05$, $p(X) = 0.9$, $p(X | \text{not-E}) = 0.5$, and either (i) $b = 0.6$ and $d = 0$ OR (ii) $b = 0.2$ and $d = 0.2$ OR (iii) $b = 0.2$ at t_1 and $b = 0$ at t_2 (the benefit is 'consumed' before the second choice) and $d = 0.26$. Expected utilities: (i) At t_1 , $EU(\text{commit}) = (0.9)(1.6) + (0.1)(0) = 1.44$, $EU(\text{believe \& re-evaluate}) = (0.85)(1) + (0.05)(0.75) + (0.05)(0) + (0.05)(0.75) = 0.925$; at t_2 , $EU(\text{stick with commitment}) = (0.5)(1.6) + (0.5)(0) = 0.8$, $EU(\text{abandon commitment}) = 0.75$. (ii) At t_1 , $EU(\text{commit}) = (0.9)(1.2) + (0.1)(0) = 1.08$, $EU(\text{believe \& re-evaluate}) = (0.85)(1) + (0.05)(0.75) + (0.05)(0) + (0.05)(0.75) = 0.925$; at t_2 , $EU(\text{stick with commitment}) = (0.5)(1.2) + (0.5)(0) = 0.6$, $EU(\text{abandon commitment}) = 0.75 - 0.2 = 0.55$; (iii) at t_1 , $EU(\text{commit}) = 1.08$, $EU(\text{believe \& re-evaluate}) = 0.925$ (as in (ii)); at t_2 , $EU(\text{stick with commitment}) = (0.5)(1) + (0.5)(0) = 0.5$, $EU(\text{abandon commitment}) = 0.75 - 0.26 = 0.49$. Note that in (i), given that $p(X | \text{not-E})$ is fairly low, the benefits of commitment must be very high to meet the conditions ($0.5b > M -$

Thus, if there are benefits to committing now, *and* these benefits can only be captured if you follow through (or do something which imposes a cost to not following through), then it can be rational both to commit to believe ‘X’ and to follow through at a later time even if it would not have been rational to believe ‘X’ anew at the later time.

Making a commitment—letting your beliefs swing free of future evidence—has both advantages and disadvantages. The disadvantage is that you will be kept from shedding a false belief, if it is indeed false; the advantages are that you will be kept from shedding a true belief, plus whatever advantages accrue from keeping that belief at the center of your epistemic and practical life, if it is indeed true. When there is a high enough probability that negative evidence would be misleading—would lead you to stop believing ‘X’ even though X holds—then the advantages occur with high enough probability to outweigh the disadvantages.

Put differently, if after you get evidence against X, there is still a sufficiently high probability that X holds, then sticking with a belief you’ve already ‘invested’ in, for the possibility that it’s true, can outweigh switching course, for the possibility that it’s false. The benefits of the commitment lower the

0.5): high enough that for many initial values of $p(X)$, both committing to believe ‘X’ and committing to believe ‘not-X’ will be preferable to agnosticism.

Cases of rational commitment, in which counter-evidence would lead to belief in ‘not-X’ in absence of a commitment: $M = 0.75$, $p(X \& E) = 0.85$, $p(X \& \text{not-}E) = 0.02$, $p(\text{not-}X \& E) = 0.05$, $p(\text{not-}X \& \text{not-}E) = 0.08$, $p(X) = 0.87$, $p(X | \text{not-}E) = 0.2$, and either (iv) $b = 0.6$ and $d = 0.5$ OR (v) $b = 0.2$ and $d = 0.6$ OR (vi) $b = 0.2$ at t_1 and $b = 0$ at t_2 (the benefit is ‘consumed’ before the second choice) and $d = 0.61$. Expected utilities:

(iv) t_1 , $EU(\text{commit}) = (0.87)(1.6) + (0.13)(0) = 1.392$, $EU(\text{believe \& re-evaluate}) = (0.85)(1) + (0.02)(0) + (0.05)(0) + (0.08)(1) = 0.93$; at t_2 , $EU(\text{stick with commitment}) = (0.2)(1.6) + (0.8)(0) = 0.32$, $EU(\text{abandon commitment}) = (0.2)(-0.5) + (0.8)(0.5) = 0.3$; (v) t_1 , $EU(\text{commit}) = (0.87)(1.2) + (0.13)(0) = 1.044$, $EU(\text{believe \& re-evaluate}) = (0.85)(1) + (0.02)(0) + (0.05)(0) + (0.08)(1) = 0.93$; at t_2 , $EU(\text{stick with commitment}) = (0.2)(1.2) + (0.8)(0) = 0.24$, $EU(\text{abandon commitment}) = (0.2)(-0.6) + (0.8)(0.4) = 0.2$; (vi) t_1 , $EU(\text{commit}) = 1.044$, $EU(\text{believe \& re-evaluate}) = 0.93$ (as in (v)); at t_2 , $EU(\text{stick with commitment}) = (0.2)(1) + (0.8)(0) = 0.2$, $EU(\text{abandon commitment}) = (0.2)(-0.61) + (0.8)(0.39) = 0.19$. Note that in (iv) - (vi) the costs of backing out must be very high to meet the conditions, even if the initial benefits are also high; furthermore, a case in which the costs of backing out are 0 would require a very high b , which significantly reduces the cases in which it is rational to stick with the commitment to believe ‘X’ when one would have believed ‘not-X’ in the absence of the commitment.

evidential bar needed to believe ‘X’ at later times. Thus, it can be rational to *maintain* a belief in the face of doubt, that you wouldn’t adopt anew, even if those benefits only obtain if the belief is true.

One worry that arises is that these benefits of committing to a belief are not epistemic benefits. If so, then my defense of the rationality of convictions seems to be on a par with defenses that appeal to wishful thinking, i.e., to the benefits to the believer of being in a state of belief regardless of the truth. Two facts to note in response. First, the benefits mentioned only obtain if the belief is *true*: they are directly tied to the alethic status of the belief. Second, the benefits in question are benefits one gets *in virtue of* representing the world correctly. This makes the benefits I focus on here different from the benefits of wishful thinking, which obtain regardless of the truth of the belief, and obtain because it is enjoyable or beneficial in itself to be in a particular doxastic state. Still, if the reader is uncomfortable with my defense of convictions on the grounds that some of the benefits I’ve discussed are non-epistemic (e.g., the benefits of confident action), then we can restrict our focus to the benefits of confident evaluation, which are epistemic and will be enough on their own to explain the phenomena that this paper focuses on.³⁵

Although it is rational to maintain a commitment in the face of negative evidence, it will not be rational to do so in the face of evidence that is too strongly negative, or in the face of mildly negative evidence for too long a period of time. If credence is low, or if credence is middling for long enough that the benefits of commitment don’t make up for the possibility of being talked out of a bad risk, then it may be rational to become agnostic: the ‘penalty’ for a false belief, along with the probability that X does not hold, may outweigh the advantages of a committed true belief if X does hold.

³⁵ A decision-theoretic defense of convictions, especially religious convictions, will naturally put the reader in mind of Pascal’s Wager, and one might wonder how the defense here compares to Pascal’s defense of religious belief. In particular, Pascal rests his defense on an infinite benefit that one gets from believing that God exists, if this belief is in fact true. While both my defense and Pascal’s rest on an additional benefit from a belief if that belief is true, there are several differences. First, Pascal focuses on *holding* a belief, where I focus on *committing to* a belief and *remaining* committed; he is primarily concerned with a synchronic phenomenon, where I am concerned with a diachronic phenomenon. Second, while Pascal’s defense is entirely practical, mine is plausibly epistemic (especially if we focus only on the benefits of confident evaluations). Third, unlike Pascal, I am not interested in defending the rationality of religious convictions *regardless* of one’s evidence; instead, I am defending the rationality of convictions *if* one’s evidence has a certain character. Finally, unlike Pascal, I consider belief in God’s existence to be similar in nature to other cases in which one should have convictions.

We saw that faith that ‘X’ can involve adopting a conviction that ‘X’: a commitment to believe ‘X’ without even in the face of counter-evidence. We’ve now seen the circumstances under which such a conviction is rational, and why.

5. Faith and Traditions

Let us return to the familiar features of a tradition: recalcitrance, gestalt, and conversion.

The first thing to show is that it can be rational to have faith in the core assumptions of a tradition—to adopt not just a belief, but a conviction in those assumptions. (And, with respect to the more general dictates of faith, to commit to *act* on the core assumptions, even if one receives counter-evidence.) Let us take each of the three conditions for the rationality of faith separately.

The first condition is that one’s credence in the core assumptions of a tradition is antecedently high. There are two plausible lines of thought that can show that one should typically begin with high credence in the core claims of a tradition. The first line points out that a tradition is an amalgamation of many thoughtful people’s work, evidence-gathering, and thinking. If one is part of a tradition and notices that it is thoughtful, serious, and sensitive to evidence, then one has reason to adopt a high credence in its pronouncements. Moreover, if one is in this kind of community, then only ideas that have been thoroughly tested will survive. Finally, if one has not (yet) encountered a viable alternative, then one does not have evidence against the core claims of one’s own tradition. The second line of thought simply holds that one *starts* with a high and robust credence in the tradition one finds oneself a part of, because there is no other alternative to skepticism.

The second condition is that once one receives counter-evidence, this pushes one’s credence in the core assumptions of a tradition to a middling value, rather than a low one: that is to say, the counter-evidence is not conclusive. Note that this counter-evidence must tell strongly enough against the core assumptions that one wouldn’t believe the assumption ‘from nowhere’ if one had gotten the counter-evidence first—this is what will make the counter-evidence give rise to a ‘crisis.’

Consider first what we might call an ‘anomaly’ in the cases of science, religion, or morality: an experimental finding, a difference between the way the world is morally or religiously and what one would expect (e.g. suffering), a counterexample or case that challenges one’s theory. In all of these cases, these are bits of data that are challenging but not conclusive—they are generally not troublesome enough

to swamp the high credence one has in favor of the core assumptions. In addition, since one has not yet exhausted the resources for explaining them, one should assign some credence to the idea that one will eventually find an explanation.

Consider next the fact of an encounter with a different tradition, or with a peer who is a practitioner of that tradition. If one adopts a *conciliationalist* view of disagreement, according to which the presence of a disagreeing peer who is roughly one's intellectual equal should make one change one's credence in the claim disagreed about, then this encounter will typically leave one with a credence that is middling, rather than conclusively against the claim.³⁶

Finally, consider the 'crisis' caused by thinking through the matter oneself and coming to a different conclusion than one's community, as might often happen in the case of morality or religion. If we adopt a conciliationalist view of disagreement between oneself and one's community—my disagreement with my community should call my community's view into question but it should also call my own view into question—then we will also meet the middling-credence condition. If your credence in *X* was previously high but upon reasoning through the matter you conclude not-*X*, then your new credence in *X* should be middling—not high, because your own opinion counts, but not low, because the opinions of your community count, too.

Thus, typical cases of counter-evidence to the core assumptions of one's tradition meet condition (ii). But they do more than meet the condition. Indeed, the condition explains why these cases are thought of as a 'crisis.' The crisis consists in the fact that after observing the counter-evidence, you wouldn't believe 'from nowhere', given your new credence, and you are aware of this—but also that (as we saw in the previous section) it can nonetheless be rational to continue believing in the core assumptions of your tradition *given* your prior commitment to it. This is the source of the discomfort: one recognizes that one wouldn't believe but for the commitment; indeed, one might rather one wasn't committed at all.

The third condition is that there are benefits to be gained from committing to believe the core assumptions of a tradition (or continuing to act on them) if they are in fact true. Two of the reasons mentioned above for why commitment has benefits are particularly apt to obtain here. The first is that commitment to a proposition shapes reasoning, allowing you, for example, to keep track only of what follows from that proposition and to pay attention only to sources that agree with those core assumptions. These benefits

³⁶ See Buchak (2021) for a full treatment of peer disagreement.

are particularly large when the proposition forms the core assumption of a tradition, because the amount of content one has to keep track of when multiple traditions are live is extremely high. In addition, these benefits can only be obtained by imposing a cost to dropping the commitment later, namely that one has to start from scratch in thinking about what follows if the assumption is false.

The other reason that committing to the core assumptions of a tradition has benefits is that long-term, risky projects are often based on the truth of one's tradition: on the core assumptions or on what follows from them. A particular scientist's life may be devoted to working out and defending some theory. If that theory is true, the only way she can succeed in working it out is by remaining committed to it despite the vicissitudes of changing evidence. A particular Catholic may only attain the highest by becoming a nun, and living in a community with other nuns. If the central tenets of Catholicism are true, then she again may only be able to do this by remaining committed even as she experiences doubts. (Surely, if Catholicism is true, life in a convent is better than a life in and out of one, and more than the sum of the value of each day in the convent.)

Thus, the three conditions for rational faith are met, and (surprisingly!) a certain amount of recalcitrance is rationally required. *Given the benefits (if the core assumptions of a tradition are true) of committing to believe these core assumptions, it can be rationally required to retain this commitment in the face of a certain amount of counter-evidence—even counter-evidence that would make it irrational to believe the core assumptions if one did not already do so.* Although one will feel some tension in retaining the belief—one will be pressed on by the fact that if one was starting from nowhere it would not be rational to hold the belief—it is beneficial to epistemic life to retain it nonetheless.

We have just seen that for an adherent of a tradition, recalcitrance in the face of counter-evidence is not only rationally permitted, but rationally required. Let us now turn to the second feature of adherence to traditions: gestalt, or the fact that adherents from two traditions can have all their evidence in common and still draw different conclusions. Notice that two individuals, one who begins with rational faith that 'X' and the other who begins with rational faith that 'not-X', can, after their initial commitment, end up with the same credence in 'X' but still rationally retain their beliefs. Indeed, they can even come to agree on all of the relevant credences and each rationally retain a belief in her initial position. The theist's initial credences that God exists can meet conditions (i) and (ii) for faith that God exists, and so she can rationally form a conviction that God exists; and at the same time the atheist's initial credences that God does not exist can meet conditions (i) and (ii) for faith that God does not exist, and so he can rationally form a conviction that God does not exist. After they meet, it can be rational for them to both have a

middling credence—a credence that would not license belief in God’s existence or non-existence apart from their prior commitment—and yet it can also be rational for each of them to remain committed to their belief.³⁷

The risky-commitment account of faith explains gestalt not by the fact that evidence looks different to adherents of different traditions, but by the fact that even if adherents of different traditions agree on the credences licensed by the evidence, it will be rational for each adherent to retain their core assumptions. Convictions are *path-dependent*: the adherent of each tradition can maintain a belief in its core assumptions, even if their evidence for these assumptions is shared and they share credences.

Finally, let us examine conversion. The puzzle is that it looks as if a change in belief concerning the core assumptions of a tradition is different in kind from ordinary changes of belief. How can this be explained, and how can it be governed by standard epistemic norms, if it is different from ordinary belief change? In fact, the account here explains the phenomenology of conversion—that it is felt to be a response to build-up and that it is a discontinuity in one’s epistemic life—while showing that it issues from the rational principles that ordinarily govern belief in response to evidence.

When one is committed to believing the core assumptions of a tradition, one’s *credence* will change with one’s evidence in the ordinary way. However, if one’s credence in those assumptions goes below a threshold, then while it will still be rational to believe in the assumptions *given* the commitment, it would not have been rational without the commitment. Thus, tension is felt and withstood via the force of the commitment. But if one’s credence drops below an even lower threshold, or remains middling for long enough, then it will be rational to drop the belief, and, if the conditions for belief that or commitment to ‘not-X’ are met, to adopt a new belief or conviction. Once one does so, and the presence of the commitment is no longer motivating, it will feel like one had reason to drop the belief for a while, because from the commitment-free perspective, one did. Furthermore, since the previous commitments were based on a smaller body of evidence, and the larger body of evidence might support drastically

³⁷ For example, let the theist begin with evidence E and the atheist with evidence F, which they then share.

Borrowing numbers from footnote [34] and letting G stand for the proposition that God exists, let everyone agree that $p(G | E) = 0.9$; $p(G | F) = 0.1$; and $p(G | E \& F) = 0.5$. Let $M = 0.75$, and let commitment have the necessary value (e.g., $b = 0.6$, or $b = 0.2$ and $d = 0.2$). Then the theist will be rational to have faith that God exists: to commit to believing that God exists, and to maintain that commitment even when her credence in G drops to 0.5. And the atheist will be rational to have faith that God does not exist: to commit to believing that God does not exist, and to maintain that commitment even when her credence in not-G drops to 0.5.

different commitments, this kind of change will often be felt as a drastic response to a build-up—an opening up of oneself to all the evidence out there—rather than an ordinary ‘belief-shift.’

Another way to state what’s going on is that credences change in a smooth way in response to small changes in evidence, and so do commitment-free beliefs; but commitment-laden beliefs (convictions) change in a step-wise fashion, even though their dynamics are partially governed by a smoothly-changing entity (credences). So convictions have different phenomenology than belief, and they are made rational by different utility values—commitment lowers the evidential bar for belief in the committed-to claim—but they are both governed by the same norms of rationality.

A different possible case of, of course, is one in which one’s *credence* changes suddenly, so that one skips the phase of middling credence with commitment to sustain the old belief and instead moves directly to the new belief. This may be more apt to occur in religion than in science; consider, for example, Saint Paul’s conversion on the road to Damascus. In this case, changes in one’s core beliefs will also be felt as a sudden shift, here in response to a large change in evidence. Thus, in this case, dropping a conviction is also not like ordinary belief shift, but for different reasons. The key point is that ordinary belief shift is a process in which one’s beliefs shift gradually as one’s evidence shifts gradually; in the case of conversion, by contrast, one’s beliefs shift dramatically, either because a gradual shift in evidence eventually leads to dropping a commitment, or because there is a dramatic shift in evidence.

Thus, we can explain why changes in belief concerning the core assumptions of a tradition are felt to be a sudden, lurching response to the evidence, rather than a gradual, gentle change. In addition, we can explain why the phenomenon of dramatic belief shift occurs primarily in the case of core assumptions of a tradition: these are the characteristic claims about which it is beneficial to have convictions rather than mere beliefs.

6. Conclusion

The nature and rationality of faith in core claims of traditions can explain recalcitrance, gestalt, and conversion, and explain them as rational responses to one’s own epistemic situation and the evidence one is confronted with. Individuals within a tradition respond recalcitrantly to challenges to that tradition because it is practically rational to ignore counter-evidence to core assumptions when choosing what to believe and how to act; and this is so because the only way a tradition can be successful *if* correct is if counterevidence is sometimes ignored. Two individuals can share evidence and yet each be rational in

believing the core assumptions of their traditions, if they (rationally) started with different convictions. And the normal process by which an individual will rationally give up a commitment to the core assumptions of a tradition involves a slow build-up of counterevidence followed by a rejection of the core claims of the tradition and a 're-setting' of new commitments on the basis of the totality of the evidence.

Nothing in this story rests on the idea that the claim one is committed to on the basis of one's current tradition and the claim of the alternative tradition are incommensurable. For all we've said, the atheist and the Christian (for example) need not be speaking different languages but simply disagree about the facts. But nothing in this story rules out the idea of incommensurability, either.³⁸ I think these are advantages of the present account.

What familiar cases of conversion have in common is not that they involve a shift from one framework to another framework that is incommensurable (though they might, or might sometimes). And what constitutes adhering to a tradition isn't framing one's evidence in the concepts of that tradition rather than the concepts of another (although, again, it might or might sometimes). Instead, adhering to a tradition involves having faith in its core claims, even when the force of the evidence shifts; and breaking from a tradition involves eschewing these commitments.

³⁸ The idea of incommensurability may be difficult to fit into a Bayesian framework. Perhaps one can replace 'faith in the claims of a tradition' with 'faith that these are the correct concepts in which to frame my view of the world', and the points in this section will hold even if incommensurability is present: a person with such faith will remain committed to seeing the world in terms of his own concepts, despite mounting evidence that these may not be the correct ones.

BIBLIOGRAPHY

- Buchak, L. (2010), 'Instrumental Rationality, Epistemic Rationality, and Evidence-Gathering,' *Philosophical Perspectives* 24 : 85–120. doi: 10.1111/j.1520-8583.2010.00186.x
- Buchak, L. (2012), 'Can it Be Rational to Have Faith?', in Jake Chandler and Victoria S. Harrison (eds.), *Probability in the Philosophy of Religion* (pp. 225-47). Oxford: Oxford University Press.
- Buchak, L. (2014), 'Rational Faith and Justified Belief,' in Tim O'Connor and Laura Goins (eds.), *Religious Faith and Intellectual Virtue* (pp. 49-73). Oxford: Oxford University Press.
- Buchak, L. (2017), 'Faith and Steadfastness in the Face of Counter-evidence', *International Journal for Philosophy of Religion* 81(1), 113-133. doi: 10.1007/s11153-016-9609-7
- Buchak, L. (2021), 'A Faithful Response to Disagreement', *Philosophical Review*. doi: 10.1215/00318108-8809893
- Davidson, D. (1973), 'On the Very Idea of a Conceptual Scheme.' *Proceedings and Addresses of the American Philosophical Association*, 47, 5-20. doi: 10.2307/3129898
- Dorst, K. (2019), 'Lockeans Maximize Expected Accuracy,' *Mind* 128(509): 175-211. doi: 10.1093/mind/fzx028
- Easwaran, K. (2016), 'Dr Truthlove, Or: How I Learned to Stop Worrying and Love Bayesian Probabilities', *Noûs* 50(4): 816-53. doi: 10.1111/nous.12099
- Elgin, C.Z. (2010), 'Persistent Disagreement', in Richard Feldman and Ted Warfield (eds.), *Disagreement* (pp. 53-68). Oxford: Oxford University Press.
- Faulkner, P. (2019), 'The nature and rationality of conversion', *European Journal of Philosophy* 27(4): 1-16. doi: 10.1111/ejop.12472
- Feyerabend, P. (1962), 'Explanation, Reduction, and Empiricism', in H. Feigl and G. Maxwell (ed.), *Scientific Explanation, Space, and Time, (Minnesota Studies in the Philosophy of Science, Volume III)* (pp. 28-97). Minneapolis: University of Minneapolis Press.
- Feyerabend, P. (1975), *Against Method*. London: New Left Books.
- Fleisher, W. (2018), 'Rational Endorsement', *Philosophical Studies* 175: 2649-75. doi: 10.1007/s11098-017-0976-4
- Gordon, A.I. (1967), *The nature of conversion*. Boston: Beacon.
- Holton, R. (2004), 'Rational Resolve', *Philosophical Review* 113: 507-35. doi: 10.1215/00318108-113-4-507
- Holton, R. (2014), 'Intention as a Model for Belief', in Manuel Vargas and Gideon Yaffe (eds.), *Rational and Social Agency: Essays on the Philosophy of Michael Bratman* (pp. 12-37). Oxford: Oxford University Press.

- Howard-Snyder, D. (2013), 'Propositional Faith: What it is and what it is Not', *American Philosophical Quarterly* 50(4): 357-72.
- Iyadurai, J. (2010), 'The Step Model of Transformative Religious Experiences: A Phenomenological Understanding of Religious Conversions in India', *Pastoral Psychology* 60: 505-21. doi: 10.1007/s11089-010-0287-6
- Jackson, E.G. (2020), 'The relationship between belief and credence,' *Philosophy Compass* 15: 1-13. doi: 10.1111/phc3.12668
- James, W. (1896), 'The Will to Believe', reprinted (excerpt) in Joel Feinberg and Russ Shafer-Landau (eds.), *Reason and Responsibility* (pp. 101-109), 2005. Belmont, CA: Wadsworth/Thomson.
- Kitcher, P. (1990), 'The division of cognitive labor', *Journal of Philosophy* 87(1), 5–22. doi: 10.2307/2026796
- Kilbourne, B. and J.T. Richardson (1989), 'Paradigm Conflict, Types of Conversion, and Conversion Theories,' *Sociological Analysis* 50(1): 1-21. doi: 10.2307/3710915
- Kox, W., W. Meeus, and H. 't Hart (1991), 'Religious Conversion of Adolescents: Testing the Lofland and Stark Model of Religious Conversion,' *Sociological Analysis* 52(3): 227-40. doi: 10.2307/3711359
- Kraay, K. (2019), 'Elgin's community-oriented steadfastness', *Synthese*. doi: 10.1007/s11229-019-02384-6
- Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Lakatos, I. (1970). 'Falsification and the Methodology of Scientific Research Programmes'. In Imre Lakatos and Alan Musgrave (eds.), *Criticism and the Growth of Knowledge*. Cambridge: Cambridge University Press, 91– 196.
- Levi, I. (1967a), *Gambling with Truth: An Essay on Induction and the Aims of Science*. Cambridge, MA: MIT Press.
- Levi, I. (1967b), 'Information and Inference', *Synthese* 17(4): 369-91. doi: 10.1007/BF00485040
- Levi, I. (1980), *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. Cambridge, MA: MIT Press.
- Levi, I. (1991), *The Fixation of Belief and Its Undoing*. Cambridge, UK: Cambridge University Press.
- Lofland, J. and R. Stark (1965), 'Becoming a World-Saver: A Theory of Conversion to a Deviant Perspective', *American Sociological Review* 30(6): 862-75. doi: 10.2307/2090965
- Lougheed, K. (2018), 'The epistemic value of deep disagreements', *Informal Logic* 38(2): 263-92. doi: 0.22329/il.v38i2.4966
- MacIntyre, A. (1977), 'Epistemological Crises, Dramatic Narrative and the Philosophy of Science', *The Monist* 60(4): 453-72. doi: 10.5840/monist197760427

- MacIntyre, A. (1988). 'The Rationality of Traditions'. In *Whose Justice? Which Rationality?* (pp. 349-360). Notre Dame, IN: University of Notre Dame Press.
- McDowell, J. (1998). 'Might There Be External Reasons?' In *Mind, Value, and Reality* (pp. 95-111). Cambridge, Massachusetts: Harvard University Press.
- Murphy, N. (1990), *Theology in the Age of Scientific Reasoning*. Ithaca, NY: Cornell University Press.
- McKaughan, D. (2017), 'On the Value of Faith and Faithfulness', *International Journal for Philosophy of Religion* 81: 7-29. doi: 10.1007/s11153-016-9606-x
- Muldoon, R. (2013), 'Diversity and the division of cognitive labor', *Philosophy Compass* 8(2): 117-25. doi: 10.1111/phc3.12000
- Pace, M. (2017), 'The Strength of Faith and Trust', *International Journal of Philosophy of Religion* 81: 135-50. doi: 10.1007/s11153-016-9611-0
- Rambo, L.R. (1993), *Understanding Religious Conversion*. New Haven, CT: Yale University Press.
- Smith, C. (2011). "On Paradigm Revolutions – How Change Happens". In *How to Go from Being a Good Evangelical to a Committed Catholic in Ninety-Five Difficult Steps* (pp. 175-190). Eugene, OR: Cascade Books.
- Strevens, M. (2003), 'The role of the priority rule in science', *Journal of Philosophy* 100(2): 55–79. doi: 10.5840/jphil2003100224
- Taylor, C. (1985). 'What is Human Agency?'. In *Human Agency and Language: Philosophical Papers Vol. 1* (pp. 15-44). Cambridge, UK: Cambridge University Press.
- van Fraassen, B.C. (2002). 'Scientific Revolution/ Conversion as a Philosophical Problem'. *The Empirical Stance*. New Haven, CT: Yale University Press, 64–110.
- Williamson, T. (2000), *Knowledge and its Limits*. New York: Oxford University Press.