



# Belief: What is it Good for?

John MacFarlane<sup>1</sup>

© The Author(s) 2023

## Abstract

“Absolutely nothing,” say the radical Bayesians. “Simplifying decisions,” say the moderates. “Providing premises in practical reasoning,” say the epistemologists. “Coordinating with others,” say I. It is hard to see how to construct an adequate theory of rational behavior without using a graded notion of belief, such as credence. But once we have credence, what role is left for belief? After surveying some answers to this question, I will explore the idea that belief is in a different line of work altogether. Its job is not to rationalize and explain an agent’s behavior, but to track what the agent would accept as a *reason*. Although some philosophers have seen the connection between belief and reasons, they have tended to see reasons as part of a theory of rational action. This locates belief in the rationalizing and explaining business, where it must vie with credence. In contrast, I argue that reasons play no essential role in an account of individual rationality; they are important because we need to coordinate with others. Credence and belief thus answer to separate needs.

## 1 Introduction

According to a long and deep-rooted philosophical tradition, we understand an agent’s actions by seeing how they are rationalized by the agent’s beliefs and goals (desires). For example, Sam is groping around on his bedside table. Why is he doing that? Because he wants to put on his glasses, and believes that they are on the table. Specifying his desire and his belief makes sense of his groping as rational and thereby explains it.

---

This is a revised version of my keynote address at the GAP.11 conference at Humboldt University in Berlin. The paper is highly programmatic and aims only to map out a path through a very dense thicket, without offering more than a cursory justification for each of the turns taken. I am grateful to two anonymous reviewers for their helpful and challenging comments.

---

✉ John MacFarlane  
jgm@berkeley.edu

<sup>1</sup> University of California, Berkeley, CA, USA

Something like this picture has been accepted by theorists of mind across the spectrum, from functionalists to interpretationists. Despite their differences, they all agree that the fundamental job description of belief, its *raison d'être*, is to rationalize and thereby explain action.

But if that is belief's job, I think we have to concede that belief doesn't do it very well. Consider the following case. Sam is walking on a narrow board that spans a shallow ditch. Why is he doing that? Because his lunch is on the other side of the ditch, and he believes that by walking across the board, he will get there. His action is rational in light of his goals and beliefs, and thus we explain it. But now let's consider a similar case, but let the board cross a chasm one hundred meters deep. Now Sam does not cross. Why not? He still wants to get to the other side, and he still believes that by walking across the board, he will get there.<sup>1</sup> Yet, if we make the gap deep enough, it ceases to be rational for him to cross.

The reason is obvious: because the expected consequences of falling are so much worse, the tiny probability of falling now amounts to a risk that is not worth the reward (lunch). Decision theory explains this beautifully. Suppose Sam's utilities for the various possible outcomes are as follows:

Crosses successfully and gets lunch	1
Stays put and skips lunch	- 1
Falls into shallow ditch	- 1
Falls into deep chasm	- 1000

Then, if Sam's credence that he will get across the board without falling is 0.99, the expected utility of crossing is 0.98 in *Shallow* and -9.01 in *Deep*, while the expected utility of not crossing is -1 in both cases. Assuming it is rational to maximize expected utility, crossing is rational in *Shallow* but not in *Deep*. On the other hand, if Sam were much more confident that he would get across (say, credence of 0.999), then crossing would be rational in both cases.

A tempting conclusion is that the binary dichotomies *believed/not believed* and *desired/not desired* are too crude to explain the differences in rational behavior between *Deep* and *Shallow*, and that what we need to do the job belief and desire were supposed to do—explaining and rationalizing behavior—are *graded* notions of credence and utility.<sup>2</sup> I can't defend this view adequately here, so I will simply

<sup>1</sup> Let's assume that Sam is not irrationally fearful of heights; he doesn't suffer any nervous reaction that would make him more likely to fall just because the ditch was deeper. So he doesn't think he's any more likely to fall into the chasm than he was in the shallow ditch. He knows that there's some *very small* chance that he will fall—but it's the *same* chance there was when the ditch was shallow.

<sup>2</sup> One might object that, even if Sam believes, in both *Deep* and *Shallow*, that he will get across the board, there are differences in his *other* beliefs that may be relevant to the rationality of crossing. For example, in *Deep* but not *Shallow*, he believes that if he falls, he will be seriously hurt. But this belief, by itself, is not enough to explain why crossing isn't rational in *Deep*. If there were guardrails, for example, or if Sam were extremely confident in his balance, then it could be rational for him to cross even while believing that falling would cause serious harm. What we would need to explain why crossing isn't rational is a belief about the *likelihood* of falling: say, the belief that he is 1% likely to fall. On some expressivist accounts, believing that one is 1% likely to fall just amounts to having a certain credence,

assume it for now, because what interests me is the question it leads to: if the graded notions are the ones we need to explain and rationalize behavior, what do we need the non-graded notions for?

My aim here is to sketch an answer to that question. I'll argue that belief *does* have a job, but it's not in the industry of rationalizing and explaining behavior. Rather, it answers to the practice of *reason-giving*. And that is something we engage in not because we're interested in acting effectively ourselves, but because we need to act together. If I am right, belief is an attitude that can sensibly be ascribed only to social reasoners, and not, for example, to a solitary AI.

What has kept philosophers from appreciating this point is their tendency to think of rationality as responsiveness to reasons. This, I'll argue, is a mistake. If we confuse rationality with responsiveness to reasons, we cannot discern the distinct roles played by credence and belief. It is then hard to avoid the conclusion that belief is a blunt instrument that has now been replaced by a better tool, the way the slide rule has been replaced by the digital computer.

## 2 Belief and Rational Action

Before I press my point that credence and belief aren't in the same line of work at all, let's see what might be said in defense of the usefulness of a concept of belief for a theory of rationality, once a graded notion like credence has been given a central role.

A natural thought is that "believes" is just an imprecise way of attributing high credences. That doesn't make it pointless. Even though we have precise ways of attributing heights, words like "tall" are still useful in ordinary life.<sup>3</sup> Nobody would claim that we have to throw away the concept of *tall* once we learn *centimeter*. So, the theoretical usefulness of the fine-grained, gradable notion of credence doesn't preclude the *practical* usefulness of the coarse-grained, absolute notion of belief (Christensen, 2004, p. 96). Still, this analysis suggests that belief won't be *theoretically* useful; just as we don't do physics with vague concepts like *tall*, we won't do psychology with *belief* (cf. Christensen, 2004, p. 97). As Stalnaker puts the point:

---

Footnote 2 (continued)

or a credal state that meets certain constraints (Moss, 2018; Schroeder, 2018; Yalcin, 2011, 2012). In that case, our explanation in terms of beliefs would be parasitic on an explanation in terms of credences and not an alternative to it. On the other hand, if believing that one is 1% likely to fall is *not* just having a credence, then it would seem to be a state that one can only be in if one has the concepts necessary to express claims about probability. Since crossing in *Deep* can be irrational even if Sam does not possess such concepts, binary belief would still not be enough to explain the facts about rationality. For present purposes, I will simply assume that we can make sense of graded credences and utilities, and that something like the decision-theoretic story vindicates rationality verdicts that we could not vindicate using just binary belief and desire.

<sup>3</sup> Why they are useful is an interesting and, I think, underdiscussed question (MacFarlane, 2020a, pp. 666–667).

Probabilistic decision theory gives a complete account of how probability values, including high ones, ought to guide behaviour, in both the context of inquiry and the application of belief outside of this context. So what could be the point of selecting an interval near the top of the probability scale and conferring on the propositions whose probability falls in that interval the honorific title ‘believed’? (Stalnaker, 1984, p. 184)

Weatherson (2005) argues that (roughly speaking) “an agent believes that  $p$  iff conditionalising on  $p$  doesn’t change any conditional preferences over things that matter” (p. 422).<sup>4</sup> Whether that is so will depend on what matters to the agent, and also on the agent’s utilities, and is therefore context-sensitive. Harsanyi, who advocates a similar view, gives this nice example:

in various minor matters I may have acted for many years on the assumption that the house next door is legally owned by the man who lives in it and may have taken it as practically certain that this was the case, even though I had only rather inconclusive evidence for it. This may have been a reasonable policy on my part because it really mattered very little to me whether he was the legal owner or not. But suppose I am now seriously considering the possibility of buying the house from him. This will obviously mean that I cannot any longer take it for practically certain that he is the legal owner (or at least that I cannot do so without much better evidence) because if I paid him for the house on the assumption that he was the legal owner even though in fact he was not, I might lose a lot of money. (Harsanyi, 1985, p. 9)

Weatherson would say that Harsanyi counts as believing that his neighbor owns the house in ordinary contexts, but not in a context where he is considering buying the house, because in that context the small probability that the neighbor doesn’t own the house cannot safely be ignored.

If this is the right thing to say about belief, though, belief is epiphenomenal. What an agent believes, on this view, is entirely determined by the agent’s credences and utilities in the context. So, although this view allows us to make sense of our ordinary talk of belief (rather than banishing it as we have banished talk of witchcraft and hysteria), it does not give belief an independent role in explaining and rationalizing behavior.

Some philosophers have tried to give belief such an independent role. On Ross and Schroeder’s view, for example, the role of belief is to simplify decision problems so that they are tractable:

The general problem...is that for any given act, an agent will typically have nonzero credence in vastly many possible consequences of this act. And so if she were to associate a given consequence with a given act-state pair only if she were certain that the act-state pair would have this consequence, then she would need to employ a vast partition of ultrafine-grained states of nature, and

<sup>4</sup> This rough view is further refined in ways that won’t matter for us here.

the resulting computational task would be unmanageable. (Ross & Schroeder, 2014, pp. 265–6)

For example, in deciding whether to take the U-Bahn or a taxi to our destination, we may assume that the U-Bahn will cost 2 euros and take 30 minutes, while the taxi will cost 25 euros and takes 15 minutes. But this way of setting up the problem excludes a lot of real possibilities. For example, we might lose our transit ticket and have to buy another one, making the U-Bahn cost 4 euros and take 35 minutes. Or there might be a transit strike, causing us to waste 30 minutes and then pay 30 euros for a taxi. On the other hand, if we take a taxi, it might get into an accident, increasing the cost dramatically. To simplify our decision, Ross and Schroeder argue, we ignore all these possibilities. Because of our cognitive limitations, then, it is rational for us to use the “heuristic of treating as true propositions about which we are uncertain” (p. 267).<sup>5</sup> But, on pain of regress, we can’t always reason about whether it’s rational to do this. So, we must be disposed to do this automatically. We must have “a *defeasible* or *default* disposition to treat them as true in our reasoning” (p. 267). That disposition is a belief. Frankish (2009)’s view that belief is a “premissing policy” is similar, as is Harsanyi (1985)’s account of acceptance.

However, both Frankish (2009) and Ross and Schroeder (2014) concede that agents need to be ready to depart from these dispositions or policies when the situation demands it. And it seems that the agent’s judgment of when that is will have to be sensitive to the agent’s credences and utilities. For example, if one adopts the policy of treating it as true that one can get across the wooden plank, but then one finds that the ditch is much deeper than one had thought, alarm bells will ring, causing one to abandon the policy. But what makes the alarm bells ring? Presumably, one’s changed expected utility calculation. (If we wait until there’s an obvious bad consequence, like falling off the board, then it’s going to be too late.)<sup>6</sup> So we have an awkward situation, not unlike that a parent faces when allowing a teenager to drive the car for the first time. The parent relinquishes control, but not really—she stands ready to take over if anything goes wrong, and must therefore exercise the full range of capacities she exercises when actually driving. It is hard to see how this kind of “dual control” psychology could really amount to a cognitive simplification.

<sup>5</sup> Ross and Schroeder say explicitly that if we only cared about the quality of our deliberative conclusions, we should just be pure Bayesian reasoners; the heuristic is justified only because we want to strike a balance between quality of the conclusions and costliness of the reasoning (p. 274).

<sup>6</sup> Similarly, when trying to decide whether to take the bus or the train, we may simplify things by ignoring low-probability outcomes like strikes by train personnel or street demonstrations that block bus traffic. But we can’t *completely* ignore these things; if getting to our destination were a life or death matter, then we’d have to take these improbable events into account.

Nozick (1993, p. 96) suggests that we switch from being belief-driven to being credence- and utility-driven when the “stakes have escalated.” What changes, when the ditch turns into a chasm, is the badness of falling. So, maybe, what happens is that we let the beliefs drive until we perceive large changes in the utilities of the possible outcomes, and then we bring in credences and reassess the situation. However, the negative utility of the outcomes (e.g. falling) can themselves depend on credences. If one has a 99% chance of surviving the fall, then falling is less bad than if one has a 50% of surviving. It seems that we must be alive to credences in order to estimate whether there has been a significant change in utilities (or “stakes”).

There are further reasons for thinking that beliefs are not just cheap labor in a workshop managed by credences. As Lara Buchak and David Owens bring out in different ways, there are some jobs for which *only beliefs will do*.

Buchak (2014, p. 299) argues that we need full belief to ground reactive attitudes like blame, resentment, indignation, guilt, or gratitude. Owens (2013, pp. 46–7) makes a similar point about a broader range of emotions, including “regret, resentment, horror, disgust, fury, sorrow, embarrassment, disappointment, shame, as well as delight, gratitude, pleasure, and pride.”<sup>7</sup> For example, if one takes pride in being Irish, one must *believe* that one is Irish: having a high credence—being almost certain that one is Irish—is not enough.

Buchak and Owens make complementary points here about the relation of belief to credence. Buchak emphasizes cases where one has a credence that is high enough to rationalize action (e.g., bets) but nonetheless lacks full belief. For example:

**Phone theft.** You leave the conference room for a few minutes, and when you return, your phone is missing. One of the three people in the room must have stolen it. You know nothing relevant about them except their genders: one man and two women. You consult statistics and find that 98% of phone thefts are done by men. On that basis, you may form a very high credence that the thief was the man.

Would it be appropriate in this case to *resent* the man? Buchak says no (and I agree). The problem is not your lack of certainty. It’s not just that resenting the man on the basis of mere high credence opens you up to the risk of wronging him by resenting him for something he didn’t do. For we can describe cases where you believe that the man stole the phone (say, on the basis of a suspicious bulge in his jacket and his refusal to make eye contact) and appropriately resent him for it, but have lower credence than you would have had on the basis of the statistics. What is necessary for resentment, Buchak argues, is not certainty but *belief*. Indeed, without belief it doesn’t even make sense to resent the man *a little bit*—discounting, as it were, for your lack of certainty (Buchak, 2014, p. 299).

There is one point on which I would like to differ from Buchak. She argues that resenting the man for stealing your phone *normatively requires* believing that he stole your phone. I would say that it *conceptually requires* believing that he stole your phone. Once we describe you as not believing that the man stole your phone, we can’t coherently describe you as resenting the man for stealing your phone, no matter how many normative transgressions we are prepared to ascribe to you. Believing is simply built into the concept of resenting.

Owens emphasizes the converse situation, where one has a credence that is *not* high enough to rationalize actions (such as bets), but nonetheless has belief, as shown by the appropriateness of the “doxastic emotions.” For example,

<sup>7</sup> Some of these are reactive attitudes in Strawson (1962)’s sense, and go on Buchak’s list too, but others—for example, pride, delight, and sorrow, are not.

**Family shame.** You have always been told, by multiple sources, that your ancestors made their fortunes on the backs of enslaved people. You have no reason to doubt this, and you are ashamed of this fact about your family. Now, you are offered a bet in which you are paid \$1 if your ancestors were slaveholders, and pay \$100 if they were not.

If you are ashamed that your ancestors were slaveholders, then you must believe that they were. Yet Owens thinks (and I agree) that it is perfectly rational not to take the bet (your credence, while high, is not above 0.99).

These observations challenge the idea that belief is just a heuristic for simplifying the cognitive challenge of operating with credences. The puzzle is not just that belief that  $p$  is possible even in cases where it would not be rational to act as if  $p$ . After all, one should expect heuristic devices to lead one astray in some cases: the whole idea of a heuristic is to trade off accuracy against simplicity. What is puzzling is why belief, if it is just a heuristic, should be required for a whole class of attitudes and emotions that have no special connection with rational action. That would seem a bizarre coincidence. These connections suggest that belief is in a different line of work altogether.

There is another mystery for the heuristics view. Belief is not just coarser-grained than credence: it also has an entirely different correctness condition. Your credence of 0.94 that the four coins I just flipped didn't all come up heads is not shown to have been incorrect when it is revealed that they did all come up heads. Indeed, as Williamson (n.d.) observes, even a credence of 1 in  $p$  does not exclude  $p$ 's being false. (For example, we should have credence 1 that an infinite number of random coin flips will not all come up heads, but this attitude does not exclude the possibility that they will all come up heads.) If the job of belief is to simplify decisions based on credences, why should it have truth as its correctness condition? We are owed a story here, and as far as I can see, nobody has discharged the debt.

Ross and Schroeder (2014) use this point about the difference in correctness conditions between belief and credence to argue against views that *reduce* belief to facts about credence (like that of Weatherson, 2005), but it seems to me that it also tells against their own "reasoning disposition account." Surely, one can assume  $p$  in order to simplify decision problems without being committed to the truth of  $p$ , and hence without opening oneself up to a charge of "incorrectness" if  $p$  turns out to be false. For example, in deciding how to bet on a coin, one might assume that the coin will land heads or the coin will land tails. If the coin then lands on its edge, does this show that one's heuristic—one's decision to deliberate *as if* landing on edge was not a possible outcome—was *incorrect*? Faced with that accusation, one might reasonably respond that one was not ruling out this outcome; one was just proceeding *as if* it was impossible, in order to simplify calculations. If belief is a heuristic for simplifying decision problems, why should it have truth as its correctness condition?<sup>8</sup>

<sup>8</sup> Ross and Schroeder say that "the reasoning disposition account implies that if  $p$  is true, then someone who believes that  $p$  is thereby right about  $p$  in the sense that she is disposed to reason in a way that involves getting things right with respect to  $p$ , whereas if  $p$  is false then someone who believes that  $p$  is thereby wrong about whether  $p$  in the sense that she is disposed to reason in a way that involves getting things wrong with respect to  $p$ " (Ross & Schroeder, 2014, pp. 275–6). I don't understand how this is an

### 3 Belief and Reasons

If beliefs are not in the business of rationalizing and explaining behavior, what line of work *are* they in? Buchak and Owens show that they have an important connection with reactive attitudes and emotions like pride and shame. We could conclude that *this* is their job: belief must figure in our theories of any creatures that have these attitudes and emotions. This is what Buchak and Owens seem to be suggesting:

...the natural home of belief—and the domain in which we cannot eliminate belief in favor of credence—is in deontological norms. (Buchak, 2014, p. 306)

...the function of Belief (and thus the source of the authority of its norms) lies in the contribution Belief makes to our emotional lives. (Owens, 2013, p. 53)

Owens observes, rightly, that this claim about function is compatible with the fact that we have many beliefs that do not engage our emotions at all. Still, it seems unsatisfying that belief should exist solely for the sake of a small class of attitudes and emotions. It is natural to think that the role of belief in our lives is more general and more fundamental than that. Should we really think that creatures who lacked the capacity for pride, shame, blame, and resentment would have no need for belief?

It is also unsatisfying to posit a brute connection between belief and the reactive attitudes and doxastic emotions, without saying something about why there should be such a connection. Why does resentment require a *binary* (non-graded) doxastic attitude to the resented fact, and why must this attitude have *truth* as its correctness condition?

I want to suggest a more general role for belief. Beliefs are not for the sake of certain emotions. They are for the sake of *reasons*. To say that an agent believes  $p$  is to say, roughly speaking, that the agent treats  $p$  as a candidate reason. Starting from this account of the *raison d'être* of belief, I will endeavor to explain what seems impossible to explain if belief is merely a heuristic device in the workshop of credence: why truth is the correctness condition of belief, and why belief (and not merely high credence) is needed for attitudes like shame and resentment.

But first: what do I mean by reasons? To regard  $p$  as a reason for  $\phi$ ing is to take the fact that  $p$  to speak in favor of  $\phi$ ing. So, the reason-for relation has (at least) two relata: a *fact* (that  $p$ ) and an action or attitude ( $\phi$ ing). Reasons are *pro tanto* and can be *outweighed* by other reasons. For example, the fact that one's jacket is unfashionable may be a reason against wearing it, while the cold temperature may speak in favor of wearing it; whether wearing the jacket is the thing to do will depend on which of these reasons is weightier. Reasons are also *defeasible*: a fact one regards as a reason to  $\phi$  may later seem not to be a reason at all, once its status as a reason has been *undercut* by another fact. The fact that it is cold may no longer be a reason for wearing a jacket if one learns that one will be in a heated building the

---

Footnote 8 (continued)

argument for their conclusion. The whole issue is whether reasoning as if  $p$  is true when in fact it is false involves "getting things wrong" in any sense; it's not enough just to *say* that it does.



whole night. Finally, reasons are *context-relative*: the fact that a number is even will generally count against its being prime, but against a background of assumed facts that includes the number's being less than 3, it may count in favor.<sup>9</sup> I won't presuppose any detailed account of reasons in what follows, though I very much like Horty (2012)'s account of reasons as the premises of triggered defaults.

*Reasoning* is the process of producing, assessing, and criticizing reasons. So understood, reasoning has only a tenuous relation to rationality. *Rationality*, as I understand it, is a matter of the internal coherence of one's attitudes, both synchronically and diachronically. This is what Bayesian epistemology and decision theory attempt to give an account of. One can be rational without ever reasoning, and without ever regarding anything as a reason. Conversely, reasoning—even good reasoning—does not necessarily make one rational. The structure of reasons relations—relations of support and defeat that hold among a relatively small number of facts and an attitude or action—is entirely different from the structure demanded by rationality, which are global relations among all of one's attitudes. One can see this most clearly by reflecting on the “requirement of total evidence” in theories of diachronic rationality, which says that it is irrational to regard something as probable just because it is probable conditional on some proper subset of one's total credal state. In reasoning, by contrast, we are always operating with a relatively small number of premises, which never amount to our full worldview.<sup>10</sup>

What, then, am I saying is the connection between belief and reasons? Consider again the logical form of the reason relation. The first *relatum* (the reason) must be a fact. So, to take something to be a candidate reason is to take it to be a fact. Taking something to be a fact is believing it. Indeed, factuality is the basic correctness condition for beliefs: if  $p$  turns out to be false, then however probable it was and however well supported by the evidence, a belief that  $p$  was incorrect. For purposes of explanation and rationalization of behavior, we don't need an attitude with this correctness condition. But for reasoning, we do. We need to keep track of the propositions that can serve as premises in reasoning—as reasons—and this is an all-or-nothing matter.

Notice that although reasons can vary in strength or weight, uncertainty does not affect the weight of a reason at all. Suppose you believe that  $p$  and regard the fact that  $p$  as a strong reason for  $\phi$ ing. If you then cease to have full belief in  $p$ , but retain a high credence,  $p$  doesn't go from being a strong reason for  $\phi$ ing to being a weaker

<sup>9</sup> See Titelbaum (2019) for some nice arguments against an *atomist* picture of reasons, according to which facts have their status as reasons independently of context. Scanlon (2014)'s account of the logical form of reasons relations explicitly relativizes them to circumstances.

<sup>10</sup> Grice (2001, pp. 81–2) asks what could correspond to a principle of total evidence in an account of reasoning, and comes up with something a lot like the sort of default logic discussed by Horty (2012). The idea is that, if you accept a conditional “if  $A$  then  $B$ ” and you accept  $A$ , and “after due search” for a further conditional with an antecedent one accepts and a consequence incompatible with  $B$  you come up empty, then you can accept the conclusion  $B$ .

reason, in your estimation. Instead, you no longer take  $p$  to be a reason at all.<sup>11</sup> Similarly, if you already believe  $p$ , an increase in your confidence that  $p$  doesn't imply taking  $p$  to be a *stronger* reason for  $\phi$ ing. I don't think it's an accident that we see something very similar with the reactive attitudes. Buchak observes:

While reactive attitudes do come in degrees, the degree of blame I assign to a particular agent is based on the severity of the act, not on my credence that she in fact did it. If I have a 0.99 credence (and full belief) that you shoplifted a candy bar, I feel a small amount of indignation toward you, but if I have a 0.2 credence (and lack a full belief) that you stole from a hungry orphan, I withhold indignation altogether, even if the mathematical expectation of how much blame you deserve is higher in the latter case. (Buchak, 2014, p. 299)

I think that Buchak's observation can be explained by the fact that belief tracks what one will accept as a reason, and by some auxiliary premises relating reactive attitudes to reasons. The reactive attitudes are reason-involving. To resent someone for  $\phi$ ing, one must take the fact that they  $\phi$ ed to be a *reason* for having this attitude. If asked why you feel sad, you can coherently reply, "no reason that I'm aware of—I just feel sad." But if asked why you resent someone, you can't very well say, "there's no reason—I just resent him." Resentment conceptually requires a reason—and the same point holds for indignation, blame, pride, and shame. If you say, "I'm feeling proud" and I ask why, you just can't say, "No reason—I just feel proud today!" Pride is constitutively a response to something you take to be a reason. If belief tracks what one takes to be a candidate reason, then the conceptual link between belief and these attitudes is explained. The connection Buchak notices between belief and the reactive attitudes and the connection Owens notices between belief and emotions like pride and shame result from more basic connections between these attitudes and reasons, and between belief and the practice of reason-giving.

#### 4 Reasons and Rationality

If what I've been arguing so far is on the right track, the binary distinction of contents into believed/not believed is important, not as part of a theory of rationality, but as part of a practice of reason-giving. An upshot is that belief attributions have no point for creatures who do not go in for reasoning. It might make sense to attribute credences to a dog, and to take the dog to be acting rationally under conditions of uncertainty, maximizing expected utility (though these attributions might be highly indeterminate, given the limited behavioral evidence we can get for a nonlinguistic creature). But there could be no real basis for saying that the dog *believes* such and such. The dog may be a rational creature, but it does not give and ask for reasons.

<sup>11</sup> You might take the fact that it is probable that  $p$  to be a reason. In this case, you'll have a belief that it is probable that  $p$ . On such beliefs, see n., above.

Here, I'm rejecting a long philosophical tradition of thinking of rationality as the faculty for apprehending and responding to reasons.<sup>12</sup> I'm even going against etymology, which derives "rational" from *ratio*, reason. But one kind of philosophical progress is the recognition that there are two distinct things which we confused before.

At this point, though, you might well wonder how much progress has been made. We started out with a question about belief, and what use it could be if we have the finer-grained notion of credence. Belief: what is it good for? We answered this question by saying that it's important for the practice of giving and asking for reasons. But we can now ask a question of much the same sort. Reasons: what are *they* good for?

From the point of view of Bayesian epistemology, traditional epistemology's talk of reasons can look superfluous and primitive—a crude way of getting at distinctions that are more accurately described by the epistemology of credences. On the Bayesian picture, we get a *degreed* notion of "counting in favor": learning *R* can increase the probability of *H* by a little or by a lot. We can explain precisely what makes what makes *D* a defeater for *R* as a reason for *H* (relative to an evidential background *E*):

$$\Pr_E(H|R\&D) < \Pr_E(H|R)$$

But instead of just saying that *D* is a defeater, we can quantify just how much it reduces the degree to which *R* counts in favor of *H*. And where traditional epistemology only measures the impact of propositions taken to be facts (since reasons are facts), Bayesian epistemology recognizes that changes in intermediate levels of credence can have a rational impact on one's other credences, too (via Jeffrey Conditionalization).

To make matters worse, in some cases reasons seem to go against rationality. We usually take the premises of deductive arguments to be conclusive reasons for their conclusions. But then we should take ourselves to have conclusive reason to believe the conjunction of the propositions we believe. The rational degree of credence in that conjunction may be quite low. (How likely is it that none of your beliefs are in error?) So it looks as if reasons are pushing us to accept something that it would be irrational to accept.<sup>13</sup>

---

<sup>12</sup> On this I agree with Broome (2013, ch. 5; Broome, 2020) and Wedgwood (2015). For a defense of the contrary view, see Kiesewetter (2017, 2020).

<sup>13</sup> Leitgeb (2014) shows that, if the "worlds" distinguished by our propositions in a context are sufficiently coarse-grained (e.g. because the question under discussion only requires making certain distinctions), then there will be threshold  $< 1$  such that logical inferences from propositions with credence above this threshold will never result in propositions with credence below the threshold. This is an important result because it shows, contrary to what is often assumed in the literature, that closure of belief under logical operations is compatible with a Lockean thesis according to which one believes that *p* iff one's credence in *p* exceeds a certain threshold. Whether the result removes the prima facie tension between reasons and rationality is not clear.

Another way to resolve the tension is to suppose that the reasons the premises of a valid deductive argument give for accepting the conclusion can be outweighed by other reasons (and are hence not conclusive).

In addition, there is some evidence from cognitive psychology that getting people to think about their reasons leads to worse (less rational) decisions.<sup>14</sup> This is not surprising, in a way. Rationality requires a delicate balance among *all* of one's attitudes, and there are many things that can have a bearing on any one decision. Sometimes being forced to select one or two of these as "the reason" causes one to ignore the others. Overestimating the importance of reasons can also lead to discounting of conclusions that are rational but not supported by full beliefs.

So, what *are* reasons good for? I want to suggest that the primary function of reason-giving is interpersonal.<sup>15</sup> Rationality demands an equilibrium of a very large number of mental states. Presumably we can attain this, at least approximately, in our own thought, because we have subpersonal mechanisms that are sensitive to irrationality. When coordinating with others, though, everything has to be under personal-level control. There are no subpersonal mechanisms that ensure that our probability distributions match. But we *can* hope to coordinate on what reasons we have, and on what counts as a reason for what. This is feasible because the structures involved in reasoning (support, defeat, undercutting, and so on) are relatively simple and surveyable.

It may be helpful here to think of belief and reasons as *digital*, in contrast to the *analog* distinctions involved in credence and rationality. When we digitize an analog signal (for example, converting a painting into a 512×512 grid of colored pixels), we ignore some of the richness contained in that signal. What do we get in return? One reason digitizing can be useful is that it makes possible efficient and reliable communication. A digital signal can be copied exactly, whereas an analog one cannot. A digital signal will often be compressed, so that it is easily recognized, attended to, agreed upon, and remembered. The binary distinction *white flag waving or not?* carries a message in war that is not likely to be lost in transmission or misunderstood, whereas any information carried by the particular pattern of the waving is less reliably transmitted. Similarly, the decision to hire Prof. X because she is the most highly rated teacher is one that different members of a hiring committee can coordinate on, even though their overall views about X's strengths and the qualities of a good hire may differ considerably. The finite, surveyable structures involved in reasoning, and the binary distinction of believed/not-believed, can be thought of as a digitization of a mental state that serves the purpose of efficient communication and coordination.<sup>16</sup>

This is not to say that reasons cannot be deployed in solitary reasoning. Obviously, we care about reasons even when we're not explicitly coordinating with others. But this individual concern with reasons looks over its shoulder at the social. Mercier and Sperber (2011, p. 61) note that some of the results from the experimental study of reasoning seems to confirm the prediction that "when people reason on their own about one of their opinions, they are likely to do so proactively, that is,

<sup>14</sup> Some of these results are summarized in Mercier and Sperber (2011, p. 69).

<sup>15</sup> For a similar conclusion, see Mercier and Sperber (2011, 2017).

<sup>16</sup> Mercier and Sperber (2017, p. 8) also emphasize fostering coordination as a primary function of reason, as well as "the effectiveness and reliability of communication" (Mercier & Sperber, 2011, p. 60).

anticipating a dialogic context, and mostly to find arguments that support their opinion.” That is: when I’m reasoning on my own, I’m trying to figure out what I could say to back up my decision, if challenged.

What really brought the social role of reasons home to me is an experience I had while on sabbatical in France. I was trying to transfer some money with PayPal, and my request was denied. I called PayPal to ask what was wrong, and I was told that their software had “flagged” my request. “Why?” I asked. “What did I do wrong, and what can I do to fix it?” “I have no idea,” said the human I was talking to. “The system is opaque to me—a black box. Wait a while and try again.”

I am sure that the AI software PayPal is running has a high degree of accuracy in identifying risky transactions. It is sensitive to a huge volume of inputs and data about historical transactions that far outstrips what would be available to an experienced human clerk. In the old days, my transaction would have been blocked by a human clerk, and the human would have given a reason—latching on to some particularly salient anomaly or discrepancy they could use to justify their denial of my transaction, and which I could have then attempted to explain away or fix. The automated system does not give a reason, and that is disconcerting to a human interacting with it. What I realized, though, is that its failure to have a reason was not really an epistemic failing on its part. There are no grounds for thinking that the system would have increased its accuracy if it had been sensitive to reasons. If we required that it give a reason (say, to conform to an EU law), then its programmers could have added a second phase where it produces a rationalization of its conclusion, but this would be causally downstream from the conclusion itself and would not affect it. Alternatively, the programmers could have changed the algorithm so that it identifies potential reasons and then does some human-style reasoning from them, but this change would likely have *decreased* the system’s accuracy. After all, it means throwing away a huge amount of the information to which the system is currently sensitive (see Babic et al., 2021), and favoring decisions that can be justified to humans, even if they are not optimal in other ways.<sup>17</sup> So the inability to provide a reason is not an epistemic problem—or at any rate, not a problem for the accuracy and reliability of the software’s conclusions. But it is a *social* problem. In the case at hand, it made it impossible for me to coordinate with PayPal—impossible to determine what I should do in order to persuade them to process my transaction.

---

<sup>17</sup> See Mercier and Sperber (2011, p. 60): “we want to explore the possibility that, even in decision making, the main function of reasoning is to produce arguments to convince others rather than to find the best decision. Thus, we predict that reasoning will drive people towards decisions for which they can argue – decisions that they can justify – even if these decisions are not optimal.” They describe a number of experiments that they take to support this prediction.

## 5 Two One-Sided Perspectives

Clarity on the distinction between rationality and reasons is essential for clarity on the distinction between credence and belief. We can see this by considering two views in the literature that fail to distinguish rationality and reasons, and that consequently fail to find a place for both belief and credence.

The first one-sided view is exemplified in this well-known quote from Richard Jeffrey:

I am inclined to think Ramsey sucked the marrow out of the ordinary notion [of belief], and used it to nourish a more adequate view. But maybe there is more there, of value. I hope so. Show me; I have not seen it at all clearly, but it may be there for all that. (Jeffrey, 1970, p. 172)

Jeffrey is looking for a role for belief in a theory of individual rational decision. He seems to assume that reasons and reasoning, if they have any role to play, would have to be part of this. As a result, he cannot find anything belief can do that credence cannot.

But there is also a one-sided view on the other side: one that appreciates the important role for beliefs in the practice of reason-giving, but tries to understand rationality in terms of this, and thus fails to see a role for *credences* in a theory of rational decision. In *Knowledge and Its Limits*, Timothy Williamson writes:

What is the difference between believing  $p$  outright and assigning  $p$  a high subjective probability? Intuitively, one believes  $p$  outright when one is willing to use  $p$  as a premise in practical reasoning. (Williamson, 2000, p. 99)

That looks a lot like the thesis I've been arguing for (though I would expunge the qualification "practical"). But what it is missing is the realization that reasoning and rationality come apart, and that the primary role of reasoning is interpersonal. For Williamson, rationality requires that one conform one's attitudes to one's evidence—that is, to what one knows, or what one considers to be one's reasons. (If you look in the index of Williamson's book under *reasons*, you'll see just one line: "see rationality.")

The conflation of reasoning and rationality is particularly clear in John Hawthorne and Jason Stanley's development of this Williamsonian theme in their paper "Knowledge and Action" (Hawthorne & Stanley, 2008). Hawthorne and Stanley argue for what they call the *Action-Knowledge Principle*, which says:

Treat the proposition that  $p$  as a reason for acting only if you know that  $p$ . [Later they seem willing to extend this to "a reason for acting or believing."]

One might ask: do this, *or else what?* Though they are not as explicit about this as one might like, their answer seems to be: “or you’ll be irrational.”<sup>18</sup> They say that “the concept of knowledge is intimately intertwined with the rationality of action” (p. 571) and describe their purpose as saying “how knowing something is related to rationally acting on it” (p. 574). So the sense of their proposal seems to be this:

the states that you take to rationalize your action (make it rational) should be states of knowledge.

Since states of knowledge must also be states of belief,<sup>19</sup> Hawthorne and Stanley are committed to denying that actions can be rationalized by credences—and they explicitly treat expected utility theory as a rival theory of rational action (p 581). Intuitively, though, it *can* be rational to act on credences in the absence of knowledge. They give this example:

Suppose I am driving to a restaurant, when I come upon a fork in the road. I think it is somewhat more likely that the restaurant is to the left than to the right. Given that these are my only options, and (say) I don’t have the opportunity to make a phone call or check a map, it is practically rational for me to take the left fork. Yet I do not *know* that the restaurant is on the left. (p. 578)

Hawthorne and Stanley agree that this action can be rational. But they can’t grant that what rationalizes it are credences. It can only be rationalized by states of knowledge. What knowledge?

Their first answer is that it is knowledge of the agent’s *epistemic probabilities* (that is, probabilities conditional on their evidence—that is, their knowledge). So, the idea is this. You know certain facts: for example, that you have a vague memory that someone told you the restaurant was on the left; that the restaurant is near a river; that the left hand road goes downhill while the right hand one goes up hill. Conditional on these categorical facts which you know, the probability that the restaurant is down the left hand road is higher than the probability that it is down the right hand road.

But this account threatens to give us a “schizophrenic” account of reasons.<sup>20</sup> Reasons are supposed to be facts—and generally speaking, they are facts about the

<sup>18</sup> In describing violations of their norm, they often use neutral words like “unacceptable” or “inappropriate” instead of “irrational.” I think that is because describing many of these cases as cases of irrationality seems quite implausible. Is it *irrational* for a sous chef to remove the cake from the oven without knowing that it is done (p. 572)? (Suppose that an overcooked cake is much worse than an undercooked one.) Is it *irrational* for parents to let their child play near a dog that they do not know to be safe (p. 573)? (What if they don’t know this dog, but know that dog bites are rare?) Is it *irrational* for Hannah to remarry, without knowing that her husband had died, when he disappears in the war and does not return for five years (p. 573)?

<sup>19</sup> Sarah Moss has argued that states of partial belief can count as knowledge (Moss, 2018). That view, if coherent, obviously makes the dialectic here more complex, and indeed, one of Moss’s motivations is to hold on to elements of a Williamsonian knowledge-first epistemology. For some comments on Moss’s view, see MacFarlane (2020b).

<sup>20</sup> Hawthorne and Stanley attribute this objection, and the label, to David Chalmers.

world, not about what you know.<sup>21</sup> If you knew that the restaurant was down the left-hand road, your reason for taking the left-hand road would be this fact about the restaurant, not a fact about your knowledge. But, on the proposed account, when you know only that it is *likely* that the restaurant is down the left-hand road, your reason is a fact about your own epistemic state. This is problematic. If a fact about your own epistemic state is a reason to turn left when you don't know that the restaurant is down the left-hand road, why shouldn't a fact about your epistemic state be a reason to turn left when you do know this?

To get around this worry, Hawthorne and Stanley suggest that the decision to take the left fork is rationalized by the known (worldly) facts that would support the judgment of high epistemic probability: for example, that the restaurant is near a river, that the left hand fork goes down hill, and that you think you remember your neighbor saying it was to the left.

Whenever we appropriately act on our knowledge of the high epistemic chance of the proposition that  $p$ , we could equally appropriately have acted on knowledge of propositions that are not about chances, viz. those propositions we know that make for a high epistemic chance of the proposition that  $p$ . (p. 584)

I think that this is a hopeless gambit. A judgment of epistemic probability depends on one's *total evidence*. If we try to replace it with a small number of facts that we might give as our *reasons* for making the epistemic probability judgment, we will inevitably leave out part of its support, and the facts that we regard as reasons will not be enough, on their own, to rationalize the action. To return to Hawthorne and Stanley's example, your knowledge

- a. that the left fork goes downhill,
- b. that the restaurant is by a river, and
- c. that you think you remember your neighbor saying that the restaurant was down the left fork

is not enough to rationalize going down the left fork. For, if you *also* knew that your neighbor bore you ill will and wanted to lead you astray, then it would be irrational to take the left fork. So, to rationalize the action we need as reasons not just a–c but also something like

- d. that you have no other evidence that counts against the restaurant's being down the left fork, or undermines the support of (a–c).

---

<sup>21</sup> Facts about what you know can of course be reasons in some cases: for example, the fact that you don't know how to fly a helicopter is a good reason not to attempt to fly one.



But (d) is again a proposition about your evidence, so we're back to the "schizophrenic" view that, in situations of uncertainty, facts about your epistemic state serve as reasons, but in situations of certainty, they do not.<sup>22</sup>

It seems to me, then, that we get into quite a muddle if we try to make sense of commonsense judgments of rationality in terms of reasons and full belief, just as we get into a muddle if we try to make sense of full belief and reasons in term of credence and rationality. My proposal is to avoid both one-sided approaches: both the one that favors belief and reasons, and the one that favors credence and rationality. To do this, we need only appreciate that beliefs are not in the business of rationalizing and explaining behavior. Instead, they track what is taken as a potential premise in a practice of reason-giving, a practice whose purpose is not improving the rationality of individual behavior, but fostering interpersonal coordination. Only in view of a clear distinction between reasons and rationality can we understand how both beliefs and credences have a role to play.

To recap, then: Belief—what is it good for? Keeping track of potential reasons. Reasons—what are they good for? Coordinating with others.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Babic, B., Gerke, S., Evgeniou, T., & Cohen, I. G. (2021). Beware explanations from AI in health care. *Science*, 373(6552), 284–286.
- Broome, J. (2013). *Rationality through reasoning*. Oxford: Blackwell.
- Broome, J. (2020). Rationality versus normativity. *Australasian Philosophical Review*, 4(4), 293–311.
- Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, 169(2), 1–27.
- Christensen, D. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford: Oxford University Press.
- Cresto, E. (2010). On reasons and epistemic rationality. *Journal of Philosophy*, 107(6), 326–330.
- Frankish, K. (2009). Partial belief and flat-out belief. In F. Huber, & C. Schmidt-Petri (Eds.), *Degrees of belief* (pp. 75–93). Dordrecht: Springer.
- Grice, P. (2001). In R. Warner (Ed.), *Aspects of reason*. Oxford: Oxford University Press.

---

<sup>22</sup> See Cresto (2010) for additional worries about Hawthorne and Stanley's reply here. One worry is that evidential probabilities are defined as probabilities conditional on the agent's total evidence, using an objective ur-prior probability distribution which may depart significantly from an agent's own personal priors. As Cresto notes, if the agent doesn't know what the ur-prior is, then there will be a gap between the facts she can state—even if she can list everything she knows—and the claim that the evidential probability of a propositional, conditional on those facts, is high.

- Harsanyi, J. C. (1985). Acceptance of empirical statements: A bayesian theory without cognitive utilities. *Theory and Decision*, 18, 1–30.
- Hawthorne, J., & Stanley, J. (2008). Knowledge and action. *Journal of Philosophy*, 105(10), 571–590.
- Horty, J. F. (2012). *Reasons as defaults*. New York: Oxford University Press.
- Jeffrey, R. (1970). Dracula meets Wolfman: Acceptance vs. partial belief. In M. Swain (Ed.), *Induction, acceptance, & rational belief* (pp. 157–185). Dordrecht: Reidel.
- Kiesewetter, B. (2017). *The normativity of rationality*. Oxford: Oxford University Press.
- Kiesewetter, B. (2020). Rationality as reasons-responsiveness. *Australasian Philosophical Review*, 4(4), 332–342.
- Leitgeb, H. (2014). The stability theory of belief. *Philosophical Review*, 123(2), 131–171.
- MacFarlane, J. (2020a). Lecture III: Indeterminacy as indecision. *Journal of Philosophy*, 117, 643–667.
- MacFarlane, J. (2020b). On probabilistic knowledge. *Res Philosophica*, 97, 97–108.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.
- Mercier, H., & Sperber, D. (2017). *The enigma of reason*. Cambridge, MA: Harvard University Press.
- Moss, S. (2018). *Probabilistic knowledge*. Oxford: Oxford University Press.
- Nozick, R. (1993). *The nature of rationality*. Princeton: Princeton University Press.
- Owens, D. (2013). Value and epistemic normativity. *Teorema: Revista Internacional de Filosofía*, 32, 35–58.
- Ross, J., & Schroeder, M. (2014). Belief, credence, and pragmatic encroachment. *Philosophy and Phenomenological Research*, 88(2), 259–288.
- Scanlon, T. M. (2014). *Being realistic about reasons*. Oxford: Oxford University Press.
- Schroeder, M. (2018). Getting perspective on objective reasons. *Ethics*, 128, 289–319.
- Stalnaker, R. (1984). *Inquiry*. Cambridge, MA: MIT Press.
- Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy*, 48, 187–211.
- Titelbaum, M. G. (2019). Reason without reasons for. *Oxford Studies in Metaethics*, 14, 189–215.
- Weatherson, B. (2005). Can we do without pragmatic encroachment? *Philosophical Perspectives*, 19(1), 417–443.
- Wedgwood, R. (2015). The pitfalls of “reasons”. *Philosophical Issues*, 25(1), 123–143.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Williamson, T. (2020). Knowledge, credence, and the strength of belief. In *Expansive epistemology: Norms, action, and the social world*. London: Routledge. <http://media.philosophy.ox.ac.uk/docs/people/williamson/Strengthofbelief.pdf>.
- Yalcin, S. (2011). Nonfactualism about epistemic modality. In A. Egan, & B. Weatherson (Eds.), *Epistemic modality* (pp. 295–332). Oxford: Oxford University Press.
- Yalcin, S. (2012). Bayesian expressivism. *Proceedings of the Aristotelian Society n s*, 112, 123–160.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.