Episteme, 15, 3 (2018) 297–312 © Cambridge University Press
doi:10.1017/epi.2018.24

RATIONAL STABILITY UNDER PRAGMATIC ENCROACHMENT

MARK SCHROEDER
maschroe@usc.edu

ABSTRACT

In this paper I will be concerned with the relationship between pragmatic encroachment and the rational instability of belief. I will be concerned to make five points: first, that some defenders of pragmatic encroachment are indeed committed to predictable rational instability of belief; second, that rational instability is indeed troublesome – particularly when it is predictable; third, that the bare thesis of pragmatic encroachment is not committed to rational instability of belief at all; fourth, that the view that Jake Ross and I have called the 'reasoning disposition' account of belief has the right structure to predict limited and stable pragmatic encroachment on the rationality of belief; and fifth and finally, that the very best cases for pragmatic encroachment are rationally stable in the right ways.

1. INTRODUCTION

According to the thesis of pragmatic encroachment in epistemology, what someone knows or is in a position to know, what it is rational for her to believe, or how confident it is rational for her to be in some proposition, can depend, in at least some cases, directly on practical features of her situation in a way that is independent of their role as evidence or as determiners of reliability. The thesis of pragmatic encroachment comes into conflict with orthodox expectations about knowledge and rational belief in at least two different ways. First, and most directly, it conflicts with the idea that only evidence, reliability, or otherwise truth-related factors can bear on knowledge or rational belief – an idea that is lent force by the familiar observation that Pascal's Wager does not make belief in God rational in the way required for knowledge. But second, in most of its guises, the thesis of pragmatic encroachment brings in its wake the idea that rational beliefs must be *unstable* – coming and going as the stakes readily shift.

I've written extensively elsewhere about the relationship between pragmatic encroachment and the orthodox idea that practical factors *can't* be relevant for rational belief or knowledge (see Schroeder 2012*a*, 2012*b*, 2013). In this paper I will be concerned with the relationship between pragmatic encroachment and the rational instability of belief. I will be concerned to make five points: first, that some defenders of pragmatic encroachment are indeed committed to predictable rational instability of belief; second, that rational instability is indeed troublesome – particularly when it is predictable; third, that the bare thesis of pragmatic encroachment is not committed to rational instability of belief at all; fourth, that the view that Jake Ross and I (Ross and Schroeder 2014) have called the 'reasoning disposition' account of belief has the right structure to predict limited and stable

pragmatic encroachment on the rationality of belief; and fifth and finally, that the very best cases for pragmatic encroachment are rationally stable in the right ways.

1.1 Setup

Before moving on, it helps to get a few clarifications out of the way. As I have stated the generic thesis of pragmatic encroachment, is the thesis that what someone knows *or* what she is in a position to know *or* what it is rational for her to believe *or* how confident it is rational for her to be can in at least some cases be affected by practical factors. But in this paper I will restrict my attention to the thesis that there is pragmatic encroachment on *rational belief*. With some exceptions (cf. Stanley 2005), most defenders of pragmatic encroachment (rightly, I think!) have not accepted pragmatic encroachment on how confident it is rational to be in any proposition (pragmatic encroachment on rational *credence*, for short). So I will set the thesis that there is pragmatic encroachment on rational credence aside, and focus just on forms of what I call *single-tier* pragmatic encroachment, which postulate pragmatic encroachment on all-or-nothing belief, but not on credence. In contrast, what I say may bear on what goes for pragmatic encroachment on knowledge, but how it bears will depend on what assumptions we make about the relationship between knowledge and rational belief. So I will set aside those questions, here.

We must also say something up front about what assumptions we are making about the nature of belief. The nature and role of all-or-nothing belief is controversial, especially in this context. Some theorists hold that rational agents like us have no such things as all-or-nothing beliefs, but only *degrees* of belief, often called *credences*, corresponding to how confident they are in different propositions (Jeffrey 2004). Others hold that credences are not psychologically real, but only binary (all-or-nothing) beliefs (especially Harman 1986; Holton 2014). Still others hold that binary belief reduces to credence or that credence reduces to binary belief.[1] And dualists accept the psychological reality of both credences and binary belief, neither of which reduces to the other (though some dualists hold that both reduce to a common core) (Ross and Schroeder 2014; Weisberg Forthcoming).

The only one of these packages of commitments that I will rule out up front is the binary belief eliminativist, who rejects the psychological reality of all-or-nothing beliefs altogether. Of course, Jake Ross and I (Ross and Schroeder 2014) have argued both directly for dualism and for the thesis that dualism is the best way to explain how there could be pragmatic encroachment on binary belief but no pragmatic encroachment on credence. And as will emerge later, in this paper I will also be arguing that dualism is the key to defending more rationally stable forms of pragmatic encroachment.

1.2 Pragmatic encroachment and rational instability

There are two traditional, common ways of motivating the thesis of pragmatic encroachment on the rationality of belief. The first goes through considering pairs of cases like the following:

> *Low Stakes*. Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. It is not important that they do

1 Compare Foley (1992), Weatherson (2005) on belief reducing to credence, and Easwaran (2016) for credences reducing to belief.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

so, as they have no impending bills. But as they drive past the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Hannah remembers the bank being open on Saturday morning a few weeks ago, so she says, 'Fortunately, it will be open tomorrow, so we can just come back.' In fact, Hannah is right – the bank will be open on Saturday.

High Stakes. Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. Since their mortgage payment is due on Sunday, they have very little in their account, and they are on the brink of foreclosure, it is very important that they deposit their paychecks by Saturday. But as they drive past the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Hannah remembers the bank being open on Saturday morning a few weeks ago, so she says, 'Fortunately, it will be open tomorrow, so we can just come back.' In fact, Hannah is right – the bank will be open on Saturday.[2]

If you judge intuitively that it is rational for Hannah to believe in *Low Stakes* that the bank will be open on Saturday, but not rational in *High Stakes*, then we can lead you to the thesis of pragmatic encroachment on rational belief by noting that the only relevant difference between the two cases is the cost of being wrong. Since the cost of being wrong is evidentially independent of whether the bank will be open on Saturday, it is not a truth-related factor. But it is an obviously practical factor. So this is a case in which a practical factor – the cost of being wrong – can affect what it is rational for Hannah to believe.

But now we can see that this argument for pragmatic encroachment carries a commitment to instability in its wake. Although the cost of being wrong is high for Hannah on Friday – because if she believes wrongly on Friday that the bank will be open on Saturday, then she will go home and come back Saturday to find it closed, and hence fail to deposit her paycheck in time for her mortgage payment to clear – it will no longer be high on Sunday. Sunday evening, it is too late for her to act on her belief that the bank was open the day before.

Because the costs of error can shift easily, if we motivate pragmatic encroachment by contrasting pairs of individuals who differ only in the cost of error, it is hard to avoid the conclusion that there will also be pairs of cases involving the same individual at different times – where shifts in the costs of error across times drive shifts in what it is rational for her to believe about the same proposition at different times. This loose relationship between case-based arguments for pragmatic encroachment and rational instability is made even tighter by the other most prominent source of motivation for pragmatic encroachment, based on what have come to be known as *knowledge-action* principles.

Knowledge-action principles can be formulated in different ways, but they all postulate a tight connection between knowledge and the basis of rational action. For example, according to Hawthorne and Stanley (2008), if you know that p, then p is available to you as a reason for action. Or according to Fantl and McGrath (2009), if you know that p, then it is rational for you to act as if p. Proponents of arguments for pragmatic encroachment based on knowledge-action principles note that what it is rational to do

2 These two cases, taken from Schroeder (2012*a*), are adapted in the first instance from Stanley (2005). In Stanley's original cases, the subjects self-ascribe knowledge, which creates a complication for whether this is best interpreted in a contextualist framework. I've eliminated this complication from my version of the cases. Similar cases can be found in DeRose (2009), Fantl and McGrath (2002), and many other places.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

in any situation is affected not just by the probabilities of various outcomes, but by the utilities and disutilities of those outcomes. So if knowledge-action principles provide any non-trivial constraint on knowledge, then a difference in the utilities of associated outcomes will sometimes affect whether someone knows.

But by similar reasoning, if knowledge-action principles provide any non-trivial constraint on knowledge, then shifts over time in the utilities of outcomes for actions can *also* shift whether someone knows – across time, as well as across individuals. Indeed, since at different times different actions are available, which will lead to different possible outcomes, knowledge-action principles can lead to extremely variable knowledge even without any underlying shifts in utilities – simply because different outcomes are available at different times.

It is no wonder, then, that proponents and critics of the thesis of pragmatic encroachment on rational belief have generally assumed that it is a part of the package of pragmatic encroachment that rational belief may be unstable – shifting to come and go as the stakes rise and fall.

2. RATIONAL INSTABILITY OF CREDENCES

Unfortunately, it is a problem for the thesis of pragmatic encroachment, if it has this consequence. That is because it is particularly bad for belief to be unstable – and bad in a way that calls into question how it could be *rational* to be unstable. But if it is irrational to be unstable, then there is no rational instability. To see what is so bad about unstable belief, we can warm up by examining one of the reasons to reject pragmatic encroachment on *credence* – the Dutch Book argument.

A Dutch Book argument works by constructing a set of bets that it would be rational for any agent who satisfies some constraint to accept, which are together guaranteed to generate a loss. Since satisfying this constraint opens you up to a guaranteed loss, this amounts to an argument that there is something irrational about satisfying this constraint. There are Dutch Book arguments for many synchronic requirements on rational credence, and also diachronic Dutch Book arguments for diachronic requirements on rational credence, including that they update by *conditionalization*. But to shift your credence in response to a shift in the stakes that is independent of the evidence is to violate conditionalization. So there are simple Dutch Book arguments against such shifts. For example, suppose that Hannah's credence on Friday (before the bank closes) that the bank will be open on Saturday is n, and that the lower stakes after the bank closes and so the choice between staying in line or coming back on Saturday is no longer live allow her to increase her credence by ε. Then it will be rational for Hannah to pay $\$1\text{-}n$ before the bank closes for a bet that pays out \$1 if the bank is not open on Saturday and \$0 otherwise, and to pay $\$n+\varepsilon$ after the bank closes for a bet that pays out \$1 if the bank is open on Saturday and \$0 otherwise. But this gives Hannah a guaranteed loss of .

Now, some believe – for the very same sorts of orthodox reasons that pragmatic encroachment is itself rejected – that Dutch Book arguments are the wrong sort of thing to justify rational constraints on belief – or at least, to justify the kinds of rational constraints – *epistemically* rational constraints – that we care about in epistemology.[3]

3 Compare discussion in Hájek (2005).

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

I'll return to this thought by the end of the paper. But being Dutch Book-able is certainly an awkward way to be. And if Hannah is at all reflective, then she will know that she is Dutch Book-able in this way, and hence freely be willing to rate the first bet as fair even while she fully expects that in the future she will soon rationally rate the second bet as fair, as well, which it is hard to see as coherent.

In any case, heightened stakes don't create a higher risk of error with respect to high credence, because rational decision-making on the basis of credence uses credences only to weight utilities. So when the disutility of going home now and coming back on Saturday, conditional on the bank being closed on Saturday, goes up, there are no extra costs associated with a high credence that the bank will be open on Saturday, because it doesn't carry with it any undue risk of incurring this high disutility. On the contrary, if the disutility is high enough, then Hannah will accept the sure cost of standing in line as a better bet than the risk of having her mortgage payment bounce.

So the direct justifications for pragmatic encroachment don't extend to pragmatic encroachment on credences, *and* the thesis of pragmatic encroachment on credences is subject to direct objections – precisely because it leads to instability in credences. For these sorts of reasons, we should reject the thesis of pragmatic encroachment on credence.

2.1 *Rational instability of belief*

Now that we've seen how problems can arise for the rational instability of credences, we are in a position to consider how they extend to belief. Reed (2012) argues that agents whose all-or-nothing beliefs shift in response to pragmatic factors can also be Dutch Booked, and so rationally unstable belief is problematic in the same way as rationally unstable credence. To put his point at its highest level of generality, if all-or-nothing belief plays any independent role in decision-making, then what choices you should be willing to make when you have a belief are different from those that you are willing to make when you do not have that belief, and this is the ammunition that we need to exploit in order to construct Dutch Books.

For example, here is a simple example of how such a Dutch Book might go.[4] Sam is riding a Ferris wheel at the carnival, and she is secured in her seat by a seatbelt. While she is at the bottom of the wheel, it is rational for her to act as if the seatbelt will hold her weight if she leans out, because the worst thing that can happen is that she would fall four feet. But when she is at the top of the wheel, it is not rational for her to act as if the seatbelt will hold her weight if she leans out, because if that is false, then she could fall to her death. Sam never tests the seatbelt by leaning out, so she never collects any new evidence about the security of her seatbelt, but she rides the Ferris wheel for five full circuits. Finally, let us suppose, since we have already seen how to construct a Dutch Book for Sam if her credences shift, that Sam's credence that the seatbelt will hold her weight if she leans out remains at .98 throughout her ride.

While Sam is at the top of the Ferris wheel, she doesn't believe that the seatbelt will hold her weight or that it will not hold her weight. But she would be willing to pay 2

4 Reed's (2012) example is different, and turns on a case involving investment strategies, with different purchases available at different times. Rightly or wrongly, I find Reed's example less persuasive, because it's not clear to me why considerations about diversity of investments must kick in only after you make a purchase, rather than before.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

cents for a bet that pays out $1 if the seatbelt would not hold her weight and $0 otherwise, because her credence that it will hold her weight is .98. When she reaches the bottom of the Ferris wheel, in contrast, although by hypothesis her credence stays the same, her belief changes – now she believes that the seatbelt will hold her weight if she leans out. So *if* beliefs make any difference for action, this belief should make a difference, here, and she should be willing to pay 99 cents for a bet that pays out a dollar if the seatbelt will hold her weight and $0 otherwise, since she believes that this is a winning bet. But in combination, these two bets lead to a guaranteed loss of a cent.

But this result has limited significance. And that is because not all pragmatic encroachers accept that binary beliefs actually do make any difference for action. According to one prominent family of views – *pragmatic credal reductivism* – all action is rationalized by credences, and all-or-nothing belief reduces to having a credence above a threshold that depends, in part, on the stakes.[5] This view is a form of single-tier pragmatic encroachment, because on this view, the stakes do not affect which credence it is rational to have, but only whether rationally having that credence counts as rationally *believing*. And on this view, Sam's belief comes and goes as she goes around the Ferris wheel, but which actions she is committed to taking does not shift, because her credences are constant and rational action depends only on credences. So on this view, Sam is not Dutch Bookable, despite the instability of her beliefs. So the Dutch Book problem is not a perfectly general one for shifty pragmatic encroachers – just for those who think that belief can play some independent role in choice or action.

But Dutch Books are only one familiar sort of reason for thinking that credences should not shift in response to practical factors that are independent of the evidence. And once we see how reasons to think that credences should not shift in response to practical factors *can* extend to the case of binary belief, we can see that other standards applying to credences can also be extended to apply to belief. So even if shifty believers are not Dutch Bookable, they still fail by other familiar standards that we use to evaluate believers.

For example, at the bottom of the Ferris wheel, Sam knows that she will soon – fully rationally – give up her belief that her seatbelt will hold her if she leans out. But since at the bottom of the Ferris wheel, Sam believes that her seatbelt will indeed hold her, it follows that she is committed to drawing the conclusion that in revising her belief, she will be believing less of the truth. So by her own lights, it is a problematic way of governing her own beliefs in service of the goal of believing the truth.

Similarly, at the bottom of the Ferris wheel, Sam believes that she will soon rationally come to give up her belief that her seatbelt will hold her. So she violates a kind of reflection-like principle on rational belief – she is committed to something that she knows she will be rationally required to give up. In the case of credences, it is not rational to be more or less confident in something than you believe that you will be rationally required to be in the future, absent forgetting something or other rational lapses. Similarly, in the case of intention, it is not rational to intend to do something that you believe that you will soon be rationally required to give up the intention to do. When Sam is at the bottom of the Ferris wheel, she violates a very similar principle with respect

5 Weatherson (2005), Ganson (2008), and on a natural interpretation, Fantl and McGrath (2009). See Ross and Schroeder (2014) for discussion.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

to belief. This still looks like a serious objection to the rational instability of belief in cases like Sam's.

2.2 *Instability is optional*

So far, we've seen why *some* pragmatic encroachers are committed to the conclusion that rational belief will be unstable in the face of shifts in the practical stakes. This commitment is suggested by the common examples used in case-based arguments for pragmatic encroachment, and it follows from arguments drawing on knowledge-action principles. And we've briefly seen some of the reasons why you might think that this kind of instability could be bad – at the very least, it seems to require reflective thinkers to see their own beliefs as not fully attuned to the truth, and to allow that it is rational to violate analogues of the principle of Reflection for binary belief – believing something that you know it will soon be rational for you not to believe.

But rational instability is not a mandatory feature of pragmatic encroachment. It is optional. There are two reasons why it does not follow from the bare thesis of pragmatic encroachment that it is rational for beliefs to be unstable over time. The first is that even if different beliefs are rational at different times (due only to pragmatic factors that are independent of evidence), it doesn't follow that it is rational to shift your belief over time in response to those factors, because there could be *diachronic* constraints on rational belief that forbid it. For example, even if the rational thing for Sam to believe about whether the seatbelt will hold her weight if she leans out while she is at the bottom of the Ferris wheel is that it will, but this is not rational for her to believe while at the top of the Ferris wheel, it could be that, having formed this belief while at the bottom of the Ferris wheel, it is not rational to give it up at the top, merely because she is at the top.

It turns out that the best empirical evidence about belief formation in the face of stakes suggests that actual belief formation behaves in this way. This is the phenomenon that Arie Kruglanski (2004) calls "seizing and freezing." If you put an agent in a situation in which it makes sense for them to form some belief, then they will continue to hold that belief even if subsequently put in a situation in which other agents do not form the belief. Having 'seized' on the information to form the belief, we then 'freeze' in that state, absent a sufficient jog to push us out of it. Any philosopher who has tried to use the same argument to convince both someone without prior commitments and someone who has already published a paper on a topic can recognize this same phenomenon anecdotally.

Abelard Podgorski (2016*a*, 2016*b*) goes further, and defends a general view on which the rationality of belief is fundamentally a constraint on the diachronic process of belief maintenance over time. On Podgorski's view, there may be synchronic facts about the level of *rational support* that some proposition has at a given time, and those synchronic facts may or may not depend on practical factors in addition to evidence, but those facts bear only on what belief it is rational to form at that time, if the agent is considering the question. It is perfectly rational, on Podgorski's view, for an agent to have a belief that it would not be rational for her to form afresh at some time, or to lack a belief that it would be rational for her to form, provided that she is not rationally required to consider or reconsider, at that time. Podgorski's view offers an explanation of why the seizing and freezing phenomenon observed by Kruglanski could be a feature of *rational* belief formation and maintenance.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

So earlier in this paper, I was being sloppy to talk about 'rational instability.' We need to distinguish between instability in what it is rational for an agent to believe, if she is considering the question afresh, and the rationality of being unstable in what one believes. If pragmatic encroachers are committed to rational instability at all, they are committed in the first instance only to instability in what it is rational to believe. But the problems of instability are all problems about an actual agent being unstable over time.

But even the conclusion that what it is rational to believe is unstable over time is optional for pragmatic encroachers. And that is because we don't need bank cases or knowledge-action arguments in order to motivate the thesis of pragmatic encroachment. Pragmatic encroachment might be true even if the only practical differences between cases that make a difference for what it is rational to believe are constituted by differences in the cross-temporal practical adequacy of some belief.

Because pragmatic encroachment has seemed so unorthodox a position in the climate of contemporary epistemology, much discussion of it has been focused on the arguments to think that it is true. So under those circumstances, it is not surprising that commitments of particular ways of arguing for pragmatic encroachment, including commitments of strong knowledge-action principles, would come to seem like parts of the package deal. But it is important to keep in view that this is what they are – commitments of certain ways of arguing for pragmatic encroachment, not commitments of the bare thesis of pragmatic encroachment, itself.

3. THE REASONING DISPOSITION ACCOUNT

In earlier work, Jake Ross and I (Ross and Schroeder 2014) defended what we called the *reasoning disposition* account of belief, in the course of our defense of a form of single-tier pragmatic encroachment. On this view, all-or-nothing belief does not reduce to facts about credence, but rather is a psychological state in its own good standing, on analogy to Michael Bratman's (1987) influential treatment of intention as its own psychological state, irreducible to combinations of belief and desire. Just as Bratman argued that intention is a solution to realistic engineering constraints on human psychology and decision making over time, we argued that belief is a solution to realistic engineering constraints on information-gathering and decision-making.

According to the reasoning disposition account of belief, the job description of the belief that p is to create a *default disposition* to rely on p in reasoning. This default is overridable, so for example, I can continue to believe that my name is 'Mark' even while rejecting a bet of a penny against my life on this proposition. I continue to be *disposed* to count on my name being 'Mark' in reasoning, but this disposition is masked in this case by a more powerful disposition to reason based on my credences when possible losses are very high. But except in extreme cases like this one that trigger other psychological mechanisms, I do rely on it. For example, in making a decision about whether to attend the party to which an invitation addressed to 'Mark' has arrived, I don't bother to insert into my decision table separate columns to capture how embarrassed I would be to be turned away, if it turns out that my name is not actually 'Mark.' Nor do I first reason about whether adding this column to my decision table is likely to make a difference, before leaving it out. I just don't consider the possibility at all.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

According to the reasoning disposition account, since the belief that *p* is itself a default disposition to count on *p* in reasoning, by default you will count on *p* in reasoning about whether to re-open the question as to whether *p*. And so if you believe that *p*, then the default answer for you to whether to re-open the question as to whether *p* will be "no, since *p*." This doesn't mean that you will never re-consider your beliefs – it just means that by *default* you will not – and indeed, that the circumstances under which you reconsider them will be a subcase of the circumstances under which the disposition involved in your belief is overridden, more generally. So it follows from the reasoning disposition account of belief that all-or-nothing beliefs are essentially ways of *overridably* but *indefinitely closing a question*.

The default non-reconsideration of belief makes beliefs, on this view, turn out to be in one additional way like intention, as Bratman conceives of it. According to Bratman, intentions are revisable, but stable by default. We don't reconsider our intention with every new piece of information that comes to light, and so sometimes we will continue to carry out a plan that it would not be rational for us to form if we were starting afresh. But the fact that intentions are stable across time in this way turns out to be, for Bratman, one of the principal advantages of intentions, because that is what allows them to play a role in enabling cross-temporal exercises in agency. Similarly, the fact that beliefs are by default stable over time allows them to play a role in reasoning that unfolds over time.

Given that beliefs, on the reasoning disposition view, are so much like intentions, it is reasonable to see them as a kind of *strategy*.[6] They are long-term, default plans for how to deal with a certain kind of situation.[7] And so it makes sense to evaluate their rationality in terms of the rationality of such strategies. Since beliefs continue by default, it is not a good strategy to adopt a belief now that you do not want to have later – the strategy must make sense at each point of time, going out into the future. This explains why beliefs are going to be rationally *stable*.

Actually, this fact about beliefs – that they are stable by default – turns out to explain *two* things about the rational stability of beliefs, corresponding to the two reasons from the last section for why pragmatic encroachment does not entail that it is rational for beliefs to be shifty over time. First, since beliefs are stable by default, a rational agent who forms a belief will continue to keep that belief even when circumstances come to change so that that belief would no longer be rational to form. So the rational believer's beliefs will be *more stable* than the circumstances under which it is rational to form that

6 I have been arguing that given the fact that the reasoning disposition account predicts that beliefs will be non-reconsidered by default, they are in one important respect analogous to intentions. But we don't need the details of the reasoning disposition account in order to maintain this analogy. It could be, for example, that the correct account of the *nature* of binary beliefs makes them out to be more like strategies, and that the dispositions to reason appealed to by the reasoning disposition account are consequences of this more general, strategy-like, nature of belief. I'll remain neutral about that, here. However, as Baron Reed has pointed out to me, beliefs are in one important respect *not* like plans. Often I form beliefs that I have little expectation of keeping far into the future – beliefs about what is going on around me at some time that does not seem particularly important or memorable, for example. But in general, I don't form plans that I expect it is likely that I will forget. This could be because beliefs are a particular kind of plan – a policy – and we do make policies that are indefinite but we expect to eventually forget about – or because though beliefs share many features with plans, this is one of their differences. But it is an important difference from normal future-directed plans.

7 Compare Justin Dallman's (2015) account of beliefs as 'credal plans.'

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

belief. This corresponds to the first reason we saw why pragmatic encroachment does not entail that it is rational for beliefs to be shifty over time.

But now we can see more – since forming a belief is taking on an ongoing commitment into the future, it will be rational to form the belief that *p* up front only if the strategy of counting on *p* in reasoning is one that is expected to bear good fruits *over time*. In the simplest case, in deciding whether to believe that *p*, you are deciding whether to *always* be disposed to count on *p* in reasoning. And this will be a good strategy just in case the benefits of simplifying your thinking will outweigh the expectation of loss from any possible future cases in which reasoning that counts on *p* would lead you to different conclusions than reasoning that does not. So the kinds of things that can make it less rational to adopt this strategy will be expectations of future cases in which the costs of error will be high enough to make a difference in what it is rational to do, but not so high as to trigger back-up mechanisms, as when I balk at accepting a bet of a penny or my life on the proposition that my name is 'Mark.'

Notice that on this picture, it matters not just whether the costs of error are high *right now*, but how high they are, spread across the entire future. So it will be much harder for a one-off case to make it irrational to form a belief now, than for a long-term set of risks, as come with the belief that I do not have a penicillin allergy, which could have a high cost of error any time I need to take an antibiotic. And to the extent that I can expect that back-up mechanisms will be triggered that will overcome my default strategy in particular cases, there is no harm in adopting the strategy. So again, if there is only one case in the future with a high cost of error, it makes sense for me to believe (for example) that the bank will be open, but plan in advance to wait in line, just for peace of mind, in the same way as I decline the bet of a penny against my life on the proposition that my name is 'Mark.' So for both of these reasons, it is very difficult for one-off costs of error to make it irrational, by themselves, to form some belief – particularly for beliefs that might possibly be relevant on future occasions.

Yet there are still some cases of stable high stakes, where the probability that it is true that *p* needs to be much higher, in order to justify the strategy of counting on *p* by default. For example, the belief that I do not have a penicillin allergy is such a case. Penicillin allergies are stable, ongoing features of our biology, which come with high risks if penicillin-based antibiotics are administered. In contrast, the costs of taking alternative antibiotics are low. So being disposed to falsely assume that you have no penicillin allergy carries a high risk of leading to a severe allergic reaction – a risk that there are low costs to avoiding. And because we can always expect more opportunities to be treated with antibiotics in the future, getting past a hospital visit without taking penicillin does not suddenly mean that we can look back on this proposition and believe it under a lower-stakes situation – the long-term expectation of risk does not substantially change after each hospital visit, unless new information about whether one has a penicillin allergy has come in. So the reasoning disposition account still has the right structure to explain why we should expect pragmatic encroachment – in cases involving stable high stakes.

3.1 The significance of stable high stakes

I've just been arguing that denials of self-attributions of penicillin allergies are stably high-stakes – they make it rational to hold out for more evidence that one has no such allergy before forming a positive belief that one does not have it, and they do so in a way that does

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

not carry risks for rational believers to end up vacillating in their beliefs over time, for these costs stay high across times. What I am now going to argue is that this – and many other cases like it – have been the best cases for pragmatic encroachment all along. A theory that gets to explain how there is pragmatic encroachment in these cases but has to give up on bank cases will not lose out on the best of what makes pragmatic encroachment an attractive view. On the contrary, it will keep the heart of it.

Consider the case of judgments of self-worth. Falsely believing yourself to be better than you are at, say, playing chess, or the piano, is a vice. But falsely believing yourself to be much worse than you really are is not a virtue – modesty, which is a virtue, is concerned with how you comport yourself, and perhaps what you *attend* to in thought, not with what you actually believe about yourself. In fact, falsely believing yourself to be much worse at something than you actually are is quite tragic. Compare the example of Virgil, who devotes seven years of his life to writing a PhD dissertation that documents the influence of Ralph Cudworth on Immanuel Kant's thinking. Years later, news comes to light on DailyNous.com of hitherto-undiscovered correspondence in which Kant asks, in 1794 "Wer ist diese Cudworth Person, die du in deinem letzten Brief erwähnt hast?"[8] If Virgil now concludes, along with other casual readers of DailyNous.com that his dissertation was worthless, this isn't modesty, but tragic. Given his personal investment, it is surely rational, rather than immodest, for him to hold out for better authentication of the letter.

Or consider a case from Marušic (2012, 2015). You are a long-time smoker and are trying to work up a resolution to quit. The empirical evidence is compelling that it takes the vast majority of smokers many attempts before they successfully quit. But if you believe that you will be unsuccessful, you won't be able to intend to quit. Marušic argues persuasively that what makes quitting even possible, in such circumstances, is that quitters *don't* believe that they will fail. He actually goes farther, and argues that it is rational for someone trying to quit smoking to believe that they will succeed, but we needn't follow him so far in order to appreciate how his case raises trouble for the idea that the only conditions governing rational belief are truth-related. The evidence that smokers fail to quit on their first attempt compares very favorably to matters about which we seem to rationally form beliefs all of the time. We even rationally expect others to fail to quit smoking on their first attempt. But if that evidence is so compelling, then it wouldn't be rationally possible to actually quit smoking. Treating the proposition that we will fail in the same way as we would treat any other proposition with respect to belief would make it impossible to rationally form an intention – much less a resolution – to quit.

Marušic's point plausibly generalizes. Jennifer Morton and Sarah Paul (MS), for example, have explored the phenomenon that psychologists call *grit* – the capacity to stick with a task in the face of evidence of failure. Many achievements worth their salt require effort and commitment extended across time in the face of difficulties and setbacks. People with greater grit manage to stick to these tasks over the full range of time required, rather than being deterred when bad news arrives, deciding not to continue because of the unlikelihood of success. But there are low points along all difficult paths – points at which evidence that one will fail predominates over recent evidence that one will make it. So arguably, grit requires *not* concluding that one will fail, in the face of this evidence. To falsely conclude that one will not be able to succeed at one of

8 "Who is this Cudworth person who you mentioned in your last letter?"

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

the low points along the path is not only to unnecessarily fail, but it is to unnecessarily fail after a great deal of expended effort. So if rationality requires us to conclude that we will fail, at these low points, then rationality requires us to miss out on the capacity for most achievements worth their salt.

All of these cases so far are first-personal, and are relatively stable cases of high-stakes propositions, because they are based in relatively enduring features of ourselves – our biology, as with the penicillin allergy case, our self-worth or ability, or our long-term projects. But stable high-stakes cases can also arise with respect to others – particularly when they are grounded in relatively enduring background facts. For example, Simon Keller (2004) and Sarah Stroud (2006) have explored the epistemic dimension of *friendship*, each arguing that it is incompatible with the demands of friendship that we apply the same evidential standards to conclusions about our friends as to similar conclusions about strangers.

For example, Billy and Stan are driving through Alabama, and are arrested for the murder of a convenience store clerk. When he is arrested, Billy confesses that they did it, and there are eye-witnesses to two young men of their description hopping into a lime-green convertible that looks just like theirs. There is even forensic evidence showing that the getaway car from the crime wore the very same tires as the car they are arrested driving. Billy and Stan's only defense is that when Billy confessed, he thought he was being arrested for stealing a can of tuna – which he had – and that there must have been another mint-green convertible with two similar-looking young men that drove by at the same time. Vinny – Billy's cousin – owes it to Billy not to respond to the evidence of guilt in the same way as local bystanders do.[9]

Both Keller and Stroud claim that epistemic rationality and the demands of friendship are simply in conflict – that it is irrational to be a good friend. But this is an unattractive conclusion. If they are right about what friendship requires, perhaps that is because friendship changes the nature of the wrong that you inflict on someone by forming false negative beliefs about them. Since the kinds of friendships that plausibly come with such epistemic demands are relatively enduring, that means that these stakes are high in a way that is relatively stable.

The role of pragmatic factors in rational belief formation about loved ones is revealed not only in the contrast between how we form beliefs about friends versus strangers, but in asymmetries in how it makes sense to form beliefs about our loved ones. For example, if my wife comes home from work and compliments me on how dinner smells, there is probably something wrong with our relationship if my first thought is to wonder whether she found a coin in a vending machine return this afternoon. Yet if she comes home and complains about how dinner smells, our relationship is in trouble if it's not part of my interpretive repertoire to consider the possibility that she has had a rough afternoon. Yet unusual compliments might be just as good of evidence that she has had an unusually good afternoon as unusual complaints are that she has had an unusually bad one.

This asymmetry, too, can be explained by the role of costs of error in belief formation. Close interpersonal relationships like this one, as Steve Bero (2017) argues, are structured around caring deeply about what each other think of one. So forming false negative beliefs about a loved one is hurtful. But whereas it is negative to interpret a spouse's attempt at a compliment as caused by a coin in a vending machine return rather than by goodwill, it is

9 These, of course, are the circumstances in the 1992 film *My Cousin Vinny*.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

positive to interpret her complaint as caused by her bad afternoon rather than by ill will. So one of these beliefs is hurtful in a way the other is not, and it is the one for which greater evidence is required. Again, the considerations that raise the stakes are the stable features of our interpersonal relationship, and so they are not liable to shift easily over time.

Even outside of the context of close interpersonal relationships, there are positive/negative asymmetries in how we form beliefs about others. For example, Tamar Gendler (2011) considers the case of the distinguished African-American historian John Hope Franklin, who was awarded the Presidential Medal of Freedom in 1995. The night before he received the award, he hosted a small party with friends at the Cosmos Club in Washington, DC, a historically white club of which he was one of the very first black members, and where most staff were black. While walking through the club, a woman calls him out, presents him with her coat check, and orders him to bring her his coat. Given that all of the club attendants were in uniform and Hope Franklin was not, it is clear that she has formed a belief that he is staff on the basis of his race alone – a statistically promising basis, given the underlying demographic statistics at the club.

Gendler argues that in cases like this one there is a conflict between epistemic rationality and avoiding implicit bias – given underlying statistical regularities in the world, many of which are directly or indirectly caused by past injustice, perfect respect for the evidence will require sometimes forming beliefs like the woman in the club. But the belief that the woman forms is racist. And I hold out hope that epistemic rationality does not require racism. If it does not, then the costs of this belief must play a role in explaining why the evidential standards are higher, for believing that a black man at a club in Washington, DC is staff. And I believe that they are – a false belief that a black man is staff not only diminishes him, but diminishes him in a way that aggravates an accumulated store of past injustice.

3.2 *How stable?*

I've now touched on all of the main points that I promised to try to make in this paper: I've shown why it is natural for pragmatic encroachers to be committed to the claim that it is rational for one's beliefs to be unstable over time, illustrated why it is natural to worry about this commitment, explained why it does not follow directly from the thesis of pragmatic encroachment itself, but only from some of its motivations, and argued that the Reasoning Disposition account of belief defended in Ross and Schroeder (2014) lends itself to pragmatic encroachment on the rationality of belief that is relatively stable. Finally, I've been briefly rehearsing a range of cases that support pragmatic encroachment but for which there is no need to think that rational belief need ever be unstable. All of these cases are I think more compelling than the traditional train and bank cases that have been appealed to so regularly in the literature on pragmatic encroachment. I'll close in this section by addressing one prominent worry.

I've said that pragmatic encroachment is not committed to any rational instability in belief in response to pragmatic factors. But the actual account that I have sketched, grounded in the Reasoning Disposition account of belief, does not quite support the idea that it is *never* appropriate to alter one's belief in response to a change in stakes – just that this will be relatively rare. And so this brings back the possibility that even though the view that I prefer only makes it rational to believe bad things about one's spouse after

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

that relationship ends, and doesn't allow Ferris wheel type shifts, that is still enough to be problematic.

The first objection that we gave to the rational instability of belief was the extended Dutch Book argument. This argument, as we noted, turned on the assumption that all-or-nothing belief can make a difference to action, but that assumption is accepted by the Reasoning Disposition view. So to construct a Dutch Book, all we need is to find someone who knows in advance that (for example), their friendship will end. Such a person will lack some negative beliefs about her friend that she may rationally come to endorse after the friendship ends, and we could use this to construct a Dutch Book. The problem with this is that if she knows in advance that her friendship will end in this way, then that is a problem with the friendship itself. Similarly, if we try to construct the Dutch Book with someone who knows in advance that she will have a close personal friendship with someone in the future, that won't work because knowing in advance that you will have such a close friendship with someone already raises the stakes for believing negative things about them now. Similar considerations apply, I believe, for Dutch Books about the other cases I've considered where the sources of the high stakes might end. So an agent who revises her beliefs in the way that I'm suggesting won't easily be able to predict in advance that she will be susceptible to a Dutch Book.[10]

Another objection that I leveled to the rational instability of belief was that it violates a natural extension of the principle of Reflection. Strictly speaking, the principle of Reflection applies to credence, not belief or intention, and it needs to be carefully qualified to be free of counterexamples. But the intuitive idea of extended Reflection for belief is that you shouldn't believe what you are very confident that in the future you won't (or shouldn't) rationally believe. Similarly, you shouldn't intend what you believe that in the future you will (or should) rationally cease to intend. Easily unstable believers run afoul of this principle.

It's important to distinguish this principle – which I accept – from a closely related principle, which I do not. The closely related principle says that you *should* believe what you are very confident you will (or should) rationally believe. Notice that the analogous principle for intention is false: it makes perfect sense to wait to make decisions that you are very confident you will make in one way rather than another, provided that the cost of making the wrong decision is high, and there is some chance that information might come in before you have to decide, which could allow you to avoid the wrong decision. In such cases, it makes perfect sense to make only a conditional decision, or even just to put the decision off. In contrast, it makes no sense at all to intend now to do something that you are very confident you will change your mind about before the time comes to act – this defeats the point of intending.

Similarly, since beliefs are much like intentions, there is nothing strange, I contend, in not yet making your mind up about something that you expect to make it up about in the future. But there *is* something strange about making your mind up about something you expect it to be rational to change your mind about in the future. This is because whereas believing something, like intending something, is being *committed* to it, but not believing something, like not intending something, is just lacking a commitment.

10 This doesn't show that there will be nothing like a Dutch Book that can happen for agents whose beliefs are structured like this, but at least it begins to show why Dutch Bookable situations will not be ordinary and systematic.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press

What it is rational to believe, given the Reasoning Disposition account of belief, depends on the long-run expectation of the costs of error, in comparison to the long-run expectation of the benefit of simplifying thought. The long-run expectation of the costs of error is not affected by how far into the future those costs are. So it's *simply not possible* to be in a low-stakes context now but expect to be in a high-stakes one in the future. That means that believers who conform their belief formation to this advice will never violate the belief Reflection principle. In contrast, it *is* possible for the expected costs of error to be mostly in the short term. And when that happens, a believer should expect that after the short term has passed, absent any new evidence coming to light, it may be best to form the belief. When this happens, it *may* be rational for a believer who follows my advice to fail to believe something that she expects that it will be rational for her to believe in the future. But as I have just argued, this is not problematic in the same way.

So although, on my view, stakes can both go up and go down, stakes never *predictably* go up – they only predictably go down, and then only rarely in a way that will change what it is rational to believe. This isn't perfect stability, but it is a lot – and it just *may* be enough.[11]

REFERENCES

Bero, S. 2017. *Responsibility and the Emotional Structure of Relationships*. PhD Dissertation, University of Southern California.
Bratman, M. 1987. *Intentions, Plans, and Practical Reason*. Chicago, IL: University of Chicago Press.
Dallman, J. 2015. *Belief as Credal Plan*. PhD dissertation, University of Southern California.
DeRose, K. 2009. *The Case for Contextualism*. Oxford: Oxford University Press.
Easwaran, K. 2016. 'Dr. Truthlove, Or How I Learned to Stop Worrying and Love Bayesian Probabilities.' *Noûs*, 50: 816–53.
Fantl, J. and McGrath, M. 2002. 'Evidence, Pragmatics, and Justification.' *Philosophical Review*, 111: 67–94.
—— 2009. *Knowledge in an Uncertain World*. Oxford: Oxford University Press.
Foley, R. 1992. 'The Epistemology of Belief and the Epistemology of Degrees of Belief.' *American Philosophical Quarterly*, 29: 111–21.
Ganson, D. 2008. 'Evidentialism and Pragmatic Constraints on Outright Belief.' *Philosophical Studies*, 139: 441–58.
Gendler, T. 2011. 'On the Epistemic Costs of Implicit Bias.' *Philosophical Studies*, 156: 33–63.
Hájek, A. 2005. 'Scotching Dutch Books?' *Philosophical Perspectives*, 19: 139–51.
Harman, G. 1986. *Change in View*. Cambridge, MA: MIT Press.
Hawthorne, J. and Stanley, J. 2008. 'Knowledge and Action.' *Journal of Philosophy*, 105: 571–90.
Holton, R. 2014. 'Intention as a Model for Belief.' In M. Vargas and G. Yaffe (eds), *Rational and Social Agency: Essays on the Philosophy of Michael Bratman*, pp. 12–37. Oxford: Oxford University Press.
Jeffrey, R. 2004. *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press.
Keller, S. 2004. 'Friendship and Belief.' *Philosophical Papers*, 33: 329–51.
Kruglanski, A. 2004. *The Psychology of Close-Mindedness*. New York, NY: Taylor and Francis.
Marušic, B. 2012. 'Belief and Difficult Action.' *Philosophers' Imprint*, 12(18): 1–30.
—— 2015. *Evidence and Agency: Norms of Belief for Promising and Resolving*. Oxford: Oxford University Press.
Morton, J. and Paul, S. MS. 'Grit.' Unpublished manuscript.

11 Special thanks to Baron Reed, Jennifer Lackey, Shyam Nair, Ram Neta, Brian Weatherson, Georgi Gardiner, Jake Ross, and the participants of the 2017 *Episteme* conference in the Galápagos.

Podgorski, A. 2016*a*. *Process-Oriented Rationality*. PhD Dissertation, University of Southern California.
—— 2016*b*. 'Dynamic Conservatism.' *Ergo*, 3: 349–76.
Reed, B. 2012. 'Resisting Encroachment.' *Philosophy and Phenomenological Research*, 85: 465–72.
Ross, J. and Schroeder, M. 2014. 'Belief, Credence, and Pragmatic Encroachment.' *Philosophy and Phenomenological Research*, 88: 259–88.
Schroeder, M. 2012*a*. 'Stakes, Withholding, and Pragmatic Encroachment on Knowledge.' *Philosophical Studies*, 160: 265–85.
—— 2012*b*. 'The Ubiquity of State-Given Reasons.' *Ethics*, 122: 457–88.
—— 2013. 'State-Given Reasons: Prevalent, if Not Ubiquitous.' *Ethics*, 124: 128–40.
Stanley, J. 2005. *Knowledge and Practical Interests*. Oxford: Oxford University Press.
Stroud, S. 2006. 'Epistemic Partiality in Friendship.' *Ethics*, 116: 498–524.
Weatherson, B. 2005. 'Can We Do Without Pragmatic Encroachment?' *Philosophical Perspectives*, 19: 417–43.
Weisberg, J. Forthcoming. 'Belief in Psyontology.' *Philosophers' Imprint*.

MARK SCHROEDER is Professor of Philosophy at the University of Southern California. He is the author of five books and over eighty articles in philosophy and is currently writing a book about the role of reasons in epistemology.

https://doi.org/10.1017/epi.2018.24 Published online by Cambridge University Press