DYNAMIC RATIONALITY AND DISPROPORTIONATE BELIEF

Wolfgang Schwarz

University of Edinburgh

1. Introduction

"A wise man", said Hume [1777/1993, part I, sec.X] "proportions his belief to the evidence". In one form or another, this *evidentialist doctrine* is widely assumed in contemporary epistemology and philosophy of science. I will argue that it is false. Rational belief need not be proportioned to the evidence. Nor, of course, does it succumb to prejudice and wishful thinking. The evidentialist doctrine is false because it clashes with compelling norms on the dynamics of rational belief.

I'm going to illustrate this clash by looking at scenarios in which an agent's evidence deteriorates over time, revealing less about the world or the agent's location than their earlier evidence. According to the evidentialist doctrine, the agent's beliefs should follow their deteriorating evidence: the agent should lose their confidence in propositions for which they used to have good evidence, without having received any contrary evidence. I will argue that the agent should instead follow a "conservative" policy and retain the earlier beliefs.

What such a scenario might look like depends, among other things, on how we understand the notion of evidence that figures in the evidentialist doctrine. I will start, in section 2, with a simple case that assumes a broadly internalist conception of evidence. In sections 3, 5, and 7, I look at cases that also target externalist accounts.

Most of my examples are somewhat strange and far-fetched. The evidentialist doctrine may well yield sensible results in ordinary situations. If so, my arguments have little practical relevance. But they matter for our philosophical theorising. Versions of the evidentialist doctrine are often endorsed in contemporary accounts of epistemic rationality (as in [Williamson, 2000], [Adler, 2002], or [Hedden, 2015]). They are widely taken for granted in discussions of uniqueness (e.g., [Kopec and Titelbaum, 2016]), peer disagreement (e.g., [Elga, 2007]), higher-order evidence (e.g., [Horowitz, 2014]), and indifference (e.g., [Greaves, 2016]). They are a defining element of Bayesian confirmation theory, in both its "objective" and its "subjective" flavour (e.g., [Howson and Urbach, 1993], [Maher, 2004]). They figure in prominent accounts

of justification (such as [Conee and Feldman, 2004], [Comesaña, 2010], and [Smith, 2017]) and in many theories of knowledge (such as [Cohen, 1988], [Lewis, 1996], and [Schaffer, 2005]). If I'm right, all these accounts need to be revised.

2. Building a Brain in a Vat

Before we turn to the first of my far-fetched scenarios, I want to briefly comment on a more common type of case that has sometimes been thought to put pressure on the evidentialist doctrine (see [Goldman, 1979, 2011], [Greco, 2011], [Kelly, 2016]).

> **Evidence Lost.** A renowned biologist tells you that frogs have three livers. Years later, you still believe that frogs have three livers, but you have forgotten how you acquired that belief.

This may seem to be a case of deteriorating evidence. When you have just talked to the biologist, you have strong evidence that frogs have three livers. Later, when you have forgotten the conversation, you no longer seem to have any evidence in support of your belief. Intuitively, however, the belief remains rational (and justified).

One problem with this kind of case is that your retained belief may well be evidence for its own truth (see, e.g., [McCain, 2015]). If you generally form beliefs about the anatomy of animals based on relevant evidence, then your belief that frogs have three livers is evidence that you once received evidence in support of this very belief. Since evidence of evidence is evidence, this suggests that you still have evidence that frogs have three livers.

Let's look at a different type of case.

> **Vat.** You decide to build an envatted duplicate of your own brain. As you set to work on Monday, you are confident that you are not a brain in a vat. On Friday, the construction is complete and your duplicate brain is switched on.

You and the newly activated brain have the same kinds of experiences, the same apparent memories, the same dispositions to accept or reject any statement.[1] It is tempting to think that you have the same evidence, and that your evidence is neutral on whether you are a brain in a vat. I claim that you should nonetheless remain confident that you are an ordinary person.

Let's work through all this more slowly.

I have stipulated that on Monday, you are rationally confident that you are not a brain in a vat. This is not crucial to my argument (as will become clear in section 3), but I it will be useful to briefly explain how it might be true, even on an internalist conception of evidence.

I assume that there is an evidential probability measure Pr, so that $\Pr(H/E)$ is the degree to which H is probable in light of E.[2] If E entails H then probability theory ensures that $\Pr(H/E)$ is 1. But $\Pr(H/E)$ can be high even if E does not entail H. Many scientific (and non-scientific) hypotheses are probable in light of our evidence without being entailed by the evidence. $\Pr(H/E)$ is greater than $\Pr(\neg H/E)$ iff $\Pr(E \wedge H)$ is greater than $\Pr(E \wedge \neg H)$. Whenever E supports H without entailing H, the evidential probability measure therefore has an a priori bias towards $E \wedge H$ and against $E \wedge \neg H$. This kind of bias may seem puzzling (compare [Das, 2022, p.115f.]), but I see no way around it. Denying it would mean denying the possibility of non-conclusive evidential support. Intuitively, the $E \wedge \neg H$ scenarios against which the measure is biased are scenarios in which the evidence E is misleading – in which things are not as one might reasonably take them to be, given the evidence E. If non-conclusive evidential support is possible, the evidential probability measure must have a bias against scenarios with misleading evidence. Brain-in-a-vat scenarios are an extreme example. For a brain in a vat, things are not at all as the evidence suggests. We should expect these scenarios to have low (prior) evidential probability.

---

1. I assume that your experiences etc. supervene on your brain state. If you think they depend on the rest of your body or your environment, change the scenario so that the relevant aspects of your body and your environment are duplicated as well. (I'll return to this version of the scenario in section 3.)
2. I assume, for simplicity, that evidential probability is unique and precise, and that an agent's total evidence can be captured by a single proposition E.

Let's grant, then, that your Monday evidence in *Vat* confers low probability on the hypothesis that you are a brain in a vat. By Friday, things have changed. The brain-in-a-vat scenario is no longer a mere possibility. It has become, in a sense, actual.

Let *E* be your total evidence on Friday, and let *H* be the self-locating hypothesis that you are not a brain in a vat.[3] The following three assumptions together entail that the evidential probability of *H* given *E* is around 1/2.

Copies. In light of *E*, it is highly probable that the world contains two live copies of your brain, one in a vat and one in an ordinary body, and that you are at one of these two locations.

Moderate Internalism. If two locations within a world are internally the same (like that of an embodied brain and its envatted duplicate), then your evidence is either compatible with both or with neither.

Self-Locating Indifference. If your evidence is compatible with two locations within the same world, then the two are equally probable in light of your evidence.

The first assumption, *Copies*, should be unproblematic. The second, *Moderate Internalism*, puts a limit on what your evidence might reveal about your place in the world.[4] Suppose two subjects within the same world have the same experiences, the same apparent memories, the same dispositions to accept sentences, and so on. *Moderate Internalism* says that if your evidence is compatible with being one of them, then it is compatible with being the other.[5] Given *Copies*, it follows that the worlds that are compatible with your Friday evidence in *Vat* predominantly contain two locations where you might be, one occupied by an ordinary person, the other by a brain in a vat. The third assumption, *Self-Locating Indifference*, implies that your evidence doesn't favour either location over the other.

Self-Locating Indifference assumes that the evidential probability measure is unbiased about self-location. The assumption is defended in [Bostrom, 2002], [Elga, 2004], and [Arntzenius and Dorr, 2017], and widely accepted in the literature. *Moderate Internalism* is also popular, with many authors endorsing much stronger forms of internalism. Both assumptions might be questioned, but let's leave this for later sections. If we accept them, we can conclude that your total evidence on Friday is neutral on whether you're an ordinary person or a brain in a vat. By the evidentialist doctrine,[6] you should become undecided about whether you are a brain in a vat.

I claim that this is wrong. You should remain confident that you are an ordinary person. Here is a quick and simple argument for why. I will give further arguments in sections 4 and 8.

---

3. Why "self-locating"? Intuitively, if you are unsure (on Friday) whether you are the newly built brain in a vat, then the object of your uncertainty isn't an objective hypothesis about the world. From an objective or third-person point of view, the scenario involves two subjects – an ordinary person, *a*, and a brain in a vat, *b*. You know, of course, that *a* is *a*, that *b* is *b*, and that *a* is not *b*. None of these objective propositions seems to capture the object of your uncertainty. I will follow the standard approach, due to [Lewis, 1979], of modelling self-locating uncertainty as uncertainty about *centred propositions*. I will often use sentences with 'you' to express these propositions. For example, the hypothesis *H*, which I expressed as '*you* are not a brain in a vat' is meant to be the centred proposition that is true at a location in a world iff the individual at that location in that world is not a brain in a vat. Most of what I will say could be translated into alternative models of self-locating belief, but I will not spell out these translations.

4. *Moderate Internalism* and *Self-Locating Indifference* are general principles about evidence and evidential probability. I have expressed them in terms of 'you' for the sake of stylistic continuity.

5. *Moderate Internalism* is entailed, for example, by any view on which (a) internally identical subjects within the same world have the same evidence, and (b) the evidential accessibility relation (that holds between one possibility and another iff the second is compatible with the evidence at the first) is reflexive and transitive. But *Moderate Internalism* does not entail either of these assumptions, and could be defended on other grounds as well.

6. There are many variations of the evidentialist doctrine. My direct target is the view, popular in Bayesian epistemology, that one's credence at any time should equal the evidential probability conditional on one's total evidence at that time. Other brands of "evidentialism" are affected by my arguments to the extent that they yield the same verdicts about my examples.

Recall that you were rationally confident on Monday that you are (and will remain) an ordinary person. Nothing you learn between Monday and Friday has any bearing on this assumption, as judged by your earlier beliefs. For example, when you switch on the brain on Friday, you learn that you now have an envatted duplicate. But you never thought that if you're going to have an envatted duplicate on Friday then it's doubtful whether you are an ordinary person. The same is true for anything else you might learn. If you had been told on Monday what you're going to learn, this would not have affected your rational confidence that you are (and will remain) an ordinary person. I say that you should not revise your beliefs in response to information that is in this sense irrelevant.

The general principle I am invoking here sometimes goes by the name of 'conservatism'[7] or 'minimal revision' or 'minimal mutilation'. Roughly, the principle says that you should not revise a belief in response to information that you rationally regarded as irrelevant to that belief before you received it. That is, if A is certain not to change its truth-value then your new credence in A after learning E (and nothing else) should equal your old credence in A whenever your old credence in A is independent of the hypothesis that E is about to be the case.

We could quibble over the precise formulation. Perhaps exceptions should be allowed for "self-verifying" or "practical" beliefs (see, e.g., [Carr, 2017], [Anscombe, 1957]), for agents who have reason to doubt their own rationality (see section 6), for agents who change epistemic perspective (e.g., [Pettigrew, 2022]) or acquire new concepts ([Steele and Stefánsson, 2021]), or for cases where there is an opportunity to improve one's beliefs without drawing on relevant evidence ([Christensen, 2000]). None of this is relevant to the *Vat* scenario.

We can even assume that in *Vat*, you don't receive any unexpected information at all. We can stipulate that whatever you learn between Monday and Friday, you knew in advance that you would learn it. The

---

7. This kind of conservatism is only loosely related to the "conservatism" discussed, for example, in [Harman, 1986], [Vahid, 2004], or [McCain, 2008].

conservative principle then reduces to the assumption that *Expected News is No News*: roughly (with the same caveats), you should not revise your beliefs in response to information of which you knew in advance that you would receive it.

3. Extending the Vat

I have argued that when you have turned on your duplicate in *Vat*, you should still be confident that you are an ordinary person, even though your evidence is neutral on this question. To show that your evidence is neutral, I relied on two assumptions about evidential probability: *Moderate Internalism* and *Self-Locating Indifference*. I mentioned that both of them could be questioned.

Williamson [2000], for example, argues that in a normal ("good") case, an agent's evidence rules out the hypothesis that they are a brain in a vat, even though there is no internal difference between the agent and their envatted counterparts. One might hold that this is true even for a case like *Vat*, where one of the envatted counterparts lives in the same world. Your Friday evidence would then rule out the vat location, but not your actual location. *Moderate Internalism* would fail.

One might also question *Self-Locating Indifference*. We've seen in the previous section that the evidential probability measure must be biased against scenarios in which the evidence is misleading. But the extent to which the evidence is misleading can vary from one location within a world to another. Concretely, one might suggest that brain-in-a-vat scenarios always have lower evidential probability than ordinary scenarios, even when the two kinds of scenarios are located in the same world. Your Friday evidence in *Vat* would then still favour being an ordinary person.

Maneuvers of this kind can't protect the evidentialist doctrine from clashing with the conservative principle.

Let's start with the second idea. Suppose the evidential probability measure favours your actual location on Friday over that of your envatted duplicate by 100 to 1. Still assuming *Moderate Internalism* (and *Copies*), the evidentialist doctrine then implies that you should become

around 99% confident that you are not the newly created brain in a vat. We still get a clash with the conservative principle because that principle may well require your credence to be even higher. Assume, for example, that you were more than 99% confident on Monday that you intend to build a brain in a vat. By the conservative principle, you should be more than 99% confident on Friday that you once had this intention, and therefore that you are not the newly created brain in a vat.[8]

What if the evidential probability measure gives *zero* probability to the vat location? In this case, consider a variant of *Vat* in which you build a copy not just of your brain, but of your entire body along with its environment. Let's say you build a giant 3D scanner/printer that can produce a perfect copy of a room and everything inside it. You set it up to produce a copy of your kitchen with you sitting at the table.[9] All the arguments from the previous section are easily adapted to this scenario, but the new scenario doesn't involve any brain in a vat.

We can go further. Consider another variant of *Vat*:

Earth. A powerful species of aliens is about to create a perfect duplicate of Earth in a distant part of the universe. Ten years later, they will turn all inhabitants of Earth – but not the inhabitants of the duplicate planet – into brains in a vat. You know all this.

---

8. If you are more than 99% confident on Monday that you intend to build a brain in a vat, and you are rational, then you are also more than 99% confident that you *either intend or have intended* to build a brain in a vat. The latter proposition is certain not to change its truth-value in the foreseeable future, so we can apply the conservative principle and infer that you should still be more than 99% confident on Friday that you either intend or have intended to build a brain in a vat. Assuming that you are certain on Friday that you are not *now* intending to build a brain in a vat, it follows that you should be more than 99% confident that you have once intended to build a brain in a vat.

9. The revised scenario resembles a well-known scenario from [Elga, 2004], in which a person called Dr Evil is informed that someone has created a perfect duplicate of him and his environment. Elga defends the evidentialist conclusion that Dr Evil should become unsure about whether he is the newly created duplicate. I think he should remain confident that he is Dr Evil.

By the conservative principle, you should remain confident that you are on Earth, with its ancient history. Ten years later (assuming you've kept track of time), you should become confident that you are a brain in a vat. As before, the evidentialist doctrine seems to disagree. This time, making the evidential probability measure biased against brain-in-a-vat scenarios would only strengthen the clash with the conservative principle.

Overall, I see no hope of preventing the clash by tweaking *Self-Locating Indifference*. The variant scenarios also show that we don't need the full strength of *Moderate Internalism*. In the limit, you and your newly built doppelganger might be perfectly alike with respect to all (extrinsic and intrinsic) qualitative properties, except for historical properties like having been alive on Monday, whose possession by an object at a time depends on the object's past. To avoid the clash between the evidentialist doctrine and the conservative principle, it's not enough to say that your evidence directly reveals certain facts about your body or your environment. Your evidence would have to directly reveal facts about your (possibly distant) past.

This is a tenable position. I am going to examine it in sections 5–7. First, I want to say more on why I think you should follow the conservative principle in cases like *Vat* and *Earth*. It's not just because I find the conservative principle intuitively plausible. The same could be said for the evidentialist doctrine. Both might be regarded as platitudes. If we had nothing to go by other than direct intuition, we might call it a tie. But there is more.

4. Four strands of dynamic incoherence

Suppose you find yourself in a *Vat*-type scenario. You follow the evidentialist doctrine and become unsure whether you are the newly created doppelganger. As a result, you no longer know that you were once a child, where you were born, where you went to school, and so on. Your belief state has become considerably less "accurate", where accuracy

measures the distance between your credence function and the truth.[10]

If you know that you follow the evidentialist doctrine, you could have foreseen this decrease in accuracy. The evidentialist doctrine thus requires your beliefs to change in a way of which you know (in advance) that it will lead you away from the truth. This is the opposite of what we should expect. If a certain move is known to lead you away from the truth, and some alternative would not, we should expect epistemic rationality to *forbid* making that move. How could it *require* the move?

Becoming unsure about whether you are the doppelganger also makes you vulnerable to an embarrassingly simple Dutch Book.[11] Suppose you create a duplicate of your entire body and its local environment. Suppose also that you want to maximize your net worth on Saturday, after the construction is finished. You aren't risk averse, and you have no other goals or aversions that would influence your betting behaviour. On Monday, when the construction begins, I offer you a deal for $0.90 that pays $1 if you once went to school and $0 otherwise. You accept. At this point, I have $0.90 and you have the deal. On Saturday morning, I offer you $0.60 for returning the deal. Equivalently, I offer you another deal for $0.40 that pays $0 if you once went to school and $1 otherwise. (I might offer the same deal to your doppelgaenger. I don't need to be able to tell you apart. I don't need to have any information that you lack.) Since you have become undecided about whether you are the newly created doppelganger, you accept. You have made a net loss of $0.30.

How is your financial misfortune relevant to your epistemic rationality? The argument goes something like this. *Premise*: An ideally rational agent whose only concern is to increase their personal net worth would not knowingly and avoidably engage in transactions that are guaranteed to reduce their net worth. If this is correct, then you (in the story) are not ideally rational. Something is wrong either with your beliefs, or with your goals, or with how these are related to your choices. But there's nothing incoherent about your goals, and the connection between your attitudes and your choices is as it ought to be: both transactions maximize expected utility. The rational fault must lie in your beliefs.[12]

I'll add one more consideration. If you become unsure whether you are the newly created doppelganger you could engage in what [Salow, 2018] calls "intentionally biased inquiry".

Imagine you have a predisposition for thyroid cancer. One day, you discover a lump in your neck that could be an early sign of thyroid cancer, but could also be a benign cyst. The possibility of having cancer terrifies you. You'd rather be confident that the lump is a cyst. If you know that you adhere to (a moderately internalist form of) the evidentialist doctrine, you could – in principle – bring about this reassuring state of mind. You could, for example, see to it that someone constructs 100 copies of you, perhaps together with your environment, with one possible difference: in all the doppelgangers the lump is a benign cyst. By the evidentialist doctrine, you should be confident, once the construction is complete, that you are one of the doppelgangers and therefore that the lump in your neck is benign. In the same way, you could (in principle) make yourself confident that your colleagues admire your intelligence, that the price of bitcoin will go up, or that climate change is a hoax, even if none of these are supported by your present evidence.

Whatever we think about the prudential merits of these schemes, they are epistemically indefensible. Indeed, aren't they precisely the kinds of moves the evidentialist doctrine is meant to guard against? If the evidence suggests that you might have cancer, you can't rationally

---

10. The use of accuracy considerations to evaluate epistemic norms is a common theme in the recent literature. See, among many others, [Pettigrew, 2016], and [Isaacs and Russell, 2023].
11. The use of Dutch Books to evaluate epistemic norms is another familiar theme in the literature. See, among many others, [Skyrms, 1993], and [Gustafsson, 2022].

12. It is not important to the argument I just gave that I, the bookie, could make a guaranteed profit. Perhaps I could not, given that I offer the second deal to both you and your doppelganger.

escape facing up to this possibility by engaging in bizarre construction projects whose feasibility has no evidential bearing on your state of health. The letter of the evidentialist doctrine here clashes with its own spirit.[13]

In sum, there are at least four reasons to think that you should retain your earlier beliefs in cases like *Vat*. Failing to do so would violate an attractive principle of conservatism. It would foreseeably decrease the accuracy of your beliefs. It would make you vulnerable to a simple Dutch Book. And it would allow you to manipulate your evidence so as to become confident in desirable propositions over which you have no control. In short, it would lead to an unpalatable kind of *dynamic incoherence*.

These considerations may not sway you. The *Vat* case, you might say, is a case of information loss, and it's no surprise that information loss leads to violation of the conservative principle, to a foreseeable decrease in accuracy, or to Dutch books of the kind I described. The "dynamic incoherence", you might say, is simply a consequence of an epistemically unfortunate situation.

To which I reply that it is not. Nothing in the *Vat* scenario forces you to become unsure whether you are an ordinary person. You could very well retain your earlier beliefs and avoid the four strands of dynamic incoherence – at the cost of having beliefs that are not proportioned to your evidence.

You may find that cost too high. You may even believe that dynamic considerations are simply irrelevant to questions of epistemic rationality. If you are committed to the evidentialist doctrine, I have little hope of

___

13. Why do I call the construction of duplicates as a form of *inquiry*? Because you would be engaged in the activity with the purpose of having the evidence it will yield. But the label isn't important. I also don't assume that one can generally evaluate practical choices from an epistemic perspective. The real premise is that epistemically rational agents are immune to blatant forms of self-deception. If you are epistemically rational and you have good reason to think that a certain hypothesis (over which you have no control) is true, you can't foreseeably make yourself believe that it is false simply by manipulating your evidence.

changing your mind. All I can hope is to convince you that your view is not an obvious platitude.

I will return to this issue in section 8, where I will explain why we should accept dynamic norms of rationality. I you want, you can skip ahead to that section. I have two pieces of unfinished business before I get there. First, I want to show that the clash between the evidentialist principle and the norms of dynamic coherence can't be avoided by adopting a strongly externalist conception of evidence. I also want to raise a question about the scope and interpretation of rational norms that bears on when an agent can be excused for violating a norm.

5.   Three Chests

In *Vat*-type scenarios, one could block the clash between the evidentialist doctrine and the norms of dynamic coherence by adopting a *historical* conception of evidence according to which an agent's evidence directly depends on their past. Such a conception may receive support from how we think about memory. In *Vat*, you and your newly created doppelganger both have apparent memories of having once gone to school, but only you have *genuine memories* of an earlier life. Your doppelganger has mere *quasi-memories*. If memories are evidence and quasi-memories are not then the evidentialist doctrine may support my verdict that you should retain your beliefs of an earlier life.

But consider a different type of case.[14]

Three Chests. It is Monday. In front of you are three chests – one red, one green, one blue. Exactly one of the chests contains a treasure. I am about to open the red chest, and you will see what's inside. If the chest is empty, nothing special will happen tonight. If the chest contains the treasure, your memories of seeing the treasure will be erased, in the following way: at midnight, I am going to put you into whatever state you would have been in if you had found the chest empty. Tomorrow, on Tuesday,

___

14. In some ways, the following scenario resembles the Sleeping Beauty problem and the "Shangri La" case from [Arntzenius, 2003].

I will open the green chest. If that chest contains the treasure, your memories of seeing the treasure will again be erased the following night (in the same way). The experiment is over on Wednesday. You know all this.

Let's assume that on Monday, you give equal credence to the three possible locations, *Red*, *Green*, and *Blue*. What then happens depends on which chest contains the treasure.

Suppose first that the treasure is in the blue chest. You are shown an empty red chest on Monday, an empty green chest on Tuesday, and nobody tinkers with your memory. On Wednesday, you have apparent memories as of finding the red and the green chest empty. Arguably, these apparent memories are genuine memories. They are veridical, and they are related to the relevant past events in the normal way.[15] If genuine memories constitute evidence, you have (on Wednesday) strong evidence that the treasure is in the blue chest. By the evidentialist doctrine, combined with the historical conception of evidence, you should be highly confident in *Blue*.

Suppose now that the treasure is in the red chest. You see it there on Monday, but these memories are erased the following night and replaced by quasi-memories of finding the red chest empty. On Tuesday, you see the empty green chest, and these memories are left intact. On Wednesday, you thus have genuine memories of seeing the empty green chest on Tuesday, and you have no (genuine) memories of what you saw on Monday. On the historical conception, your evidence is

neutral between *Blue* and *Red*, but rules our *Green*.[16] By the evidentialist doctrine, your credence in *Blue* should increase from $1/3$ to $1/2$.

The final case, where the treasure is in the green chest, is analogous to the previous case. This time, your memory evidence on Wednesday is neutral between *Blue* and *Green* and rules out *Red*. Your credence in *Blue* should again increase from $1/3$ to $1/2$.

Combined with the historical conception of evidence, the evidentialist doctrine therefore says that your credence in *Blue* should increase no matter what you observe, and no matter where the treasure is located. This is incompatible with the norms of dynamic coherence. If your credence in *Blue* increases no matter what you observe, you violate the principle of conservatism. Your beliefs foreseeably decrease in expected accuracy. You are vulnerable to a straightforward Dutch Book. (I sell you a bet on ¬*Blue* on Sunday and buy it back at a reduced price on Wednesday.) You could engage in intentionally biased inquiry. (You could, for example, instruct me on Sunday to put the treasure into the blue chest iff the lump in your neck is a benign cyst. By Wednesday, your credence in this desirable hypothesis will be certain to go up.)

The historical conception of evidence may block the clash between the evidentialist doctrine and the norms of dynamic coherence in *Vat*. It does not do so in *Three Chests*.

6. Architectural rationality

What do the norms of dynamic coherence say about *Three Chests*? They say that you should retain any information you receive. If you saw the treasure in the red chest on Monday, you should still be confident in *Red* on Wednesday. If you saw the treasure in the green chest on Tuesday,

---

15. One might hold that apparent memories are only genuine memories (or that their content only qualifies as evidence) if they could not easily have been false (see, e.g., [Weatherson, 2015]). If so, let's stipulate that the treasure could not easily have been anywhere else. Perhaps there is a law of nature, unknown to you, that makes it impossible for the treasure to be in any chest other than the blue chest. We could also consider a variant of the scenario in which you all along have good, but inconclusive reasons to suspect that the treasure is in the blue chest, so that the alternative *Red* and *Green* possibilities are not only "modally distant" and "non-normal", but also unlikely by your own lights.

16. The historical conception of evidence here invalidate the *Negative Access* principle ¬$E\phi \rightarrow E\neg E\phi$, where $E\phi$ means that the evidence entails ϕ. ¬$E(Red)$ is true, but $E\neg E(Red)$ is false. Failure of *Negative Access* is a common feature of evidential externalism, and it is known to have troublesome implications for agents who update their beliefs by conditionalizing on their evidence. See, for example, [Bronfman, 2014], [Schoenfield, 2017], [Salow, 2018], [Gallow, 2021], [Das, 2022]. We are going to see that these problems carry over to agents who obey the evidentialist doctrine.

you should remain confident in *Green*. If you found both chests empty, you should be confident in *Blue*. Could you conform to these demands, given the threat of memory erasure? If not, it may seem unfair to blame the evidentialist doctrine for recommending an update that violates dynamic coherence.

Let's have a closer look. Suppose again that the treasure is in the red chest. You see it there on Monday. At this point, you are confident in *Red*. Dynamic coherence demands that you retain this confidence in the transition from Monday to Tuesday. Could you comply with this demand, given the memory erasure at midnight?

You could, but only by having a strangely dogmatic update disposition. I've stipulated that the effect of the "memory erasure" is that your Tuesday state isn't sensitive to what you learn on Monday: you are going to be put into whatever state you would have been in if you had found the red chest empty.[17] But suppose you have a predisposition to become confident in *Red* upon falling asleep on Monday, no matter what you saw that day. The "memory erasure" then preserves your confidence in *Red*, for you would have that confidence even if you had found the red chest empty.[18]

Similar reasoning applies to Tuesday, where you learn that the green chest is empty. If you've retained your belief in *Red*, this doesn't come

---

17. I assume, for the sake of the example, that the relevant instance of Conditional Excluded Middle holds: there is a fact of the matter of what you would have done if you had found the red chest empty.

18. There are other ways of understanding memory erasure. For example, I could have merely stipulated that you will be implanted with vivid memory experiences as of finding the red chest empty. You could then easily comply with the norms of dynamic rationality. Knowing that you would have these experiences no matter what, I'd say you should dismiss them as irrelevant and hold on to your previous belief in *Red*. Alternatively, I could have stipulated that you will be implanted with a strong belief that the red chest is empty. This might still allow you to recover the truth about *Red*. Suppose, for example, that you are predisposed to become undecided about *Red* in response to seeing the treasure in the red chest. Upon awakening, you could then introspect your degree of belief in *Red* and use it to figure out what you saw on Monday, knowing that *Red* is true iff you are undecided about *Red*. My stipulation that your Tuesday state is insensitive to your Monday state blocks all such attempts to recover the truth about *Red*.

as a surprise. The norms of dynamic coherence now demand that you retain your confidence in *Red* and ¬*Green* in the transition to Wednesday. As before, you can comply with this demand, by having a disposition that would make you confident in *Red* and ¬*Green* no matter what. On Wednesday, you would still be confident in *Red*, as required by dynamic coherence.

In the same manner, suitably insensitive update dispositions would ensure that you comply with the demands of dynamic coherence if the treasure is in the green or the blue chest.

To be clear: you can't make your Wednesday state sensitive to what you learned on Monday and Tuesday. I'm not saying that you could comply with the demands of dynamic coherence by adopting different "insensitive" update dispositions depending on what you see. Suppose you form an insensitive update disposition to become confident in *Red in response to* seeing the treasure in the red chest, and that you would have formed a different disposition – to become confident in ¬*Red* – if you had found the red chest empty. The memory erasure then ensures that you wake up confident in ¬*Red* on Tuesday. You will have violated the norms of dynamic coherence. The only way to comply with dynamic coherence is to have a truly insensitive update disposition that ignores what you learn on Monday and Tuesday. If you are hard-wired to become confident that the treasure is in the red box on Monday night, no matter what you saw, you can comply with the norms of dynamic coherence – provided the treasure really is in the red box. If it is somewhere else, the norms require a different wiring.

It may help to compare this puzzling verdict with a popular externalist take on perceptual illusion. Imagine you are looking at an ordinary red wall in an ordinary setting. According to a range of externalist views (including those of [Williamson, 2000], [Goldman, 2009], and [McDowell, 2011]), your evidence entails that the wall is red. If instead you had been looking at a white wall illuminated by red light then (according to these views) your evidence would have been neutral on whether the wall is red or white. Combined with the evidentialist doctrine, these views entail that what you should believe depends on

whether you are in a "good case" or in a "bad case". In a good case, where the wall is red, you should believe that the wall is red. In a bad case, where the wall is white but looks red, you should not. This does not assume that your cognitive system is sensitive to which of the two cases you are in. The idea is not that you should have special sensory capacities that allow you to distinguish a red wall from a white wall under red light. Rather, the idea is that *what you should believe* depends on external facts about your situation. The demands of rationality might be sensitive to a feature of the world even if your cognitive system is insensitive to that feature. In *Three Chests*, your cognitive system on Wednesday is insensitive to what you have learned on Monday and Tuesday. The demands of rationality, however, might still be sensitive to what you have learned.

You may feel uneasy about this kind of dependence. I certainly do. But what's wrong with it? I don't think the problem lies with externalism. What you learn on Monday and Tuesday is, in any reasonable sense, internal to you, as a temporally extended agent. The real problem, I suspect, is that we naturally interpret epistemic norms as *architectural norms*, as (partial) construction manuals for ideal agents. On this interpretation, the norms don't just constrain what an agent actually does, but also what they do under counterfactual circumstances. So understood, the classical Bayesian norm of conditionalization, for example, requires not only that an agent's new credence happens to equal their previous credence conditional on their new evidence. It requires that the agent has a general disposition that ensures this equality, no matter what they learn.

Externalist accounts of rationality tend to be incompatible with an architectural model of rationality. One couldn't build an agent who has one kind of belief when confronting a red wall and another when confronting a white wall under red light, if their cognitive system is insensitive to the difference. Much of the intuitive resistance many people feel towards externalism may have its roots in an architectural conception of epistemic norms.

Be that as it may, what should we say about *Three Chests* if we accept (as I do) an architectural reading of dynamic norms? One option is to say that you should have an architectural disposition to retain your earlier beliefs as long as nobody intervenes and overwrites your memory. If the treasure is in the blue chest, you would then become sure of *Blue*: you would have retained the information ¬*Red* and ¬*Green*, and nobody has intervened. In the other cases, you would consequently also become sure of *Blue*. But this, one might argue, isn't your fault, as the false belief was directly inserted into your brain.

I prefer a different response. I think you shouldn't ignore the threat of memory erasure. To compensate for it, you should become indifferent between *Red*, *Green*, and *Blue* (no matter what you learn). To cut a long story short: since you know that your Wednesday beliefs are insensitive to what you learn on Monday and Tuesday, you should update your beliefs as if you aren't learning anything.[19]

Let's take stock. Standard versions of the evidentialist doctrine clash with the norms of dynamic coherence in *Vat*-type scenarios. The clash could be avoided by combining the evidentialist doctrine with a historical conception of evidence. In *Three Chests*, this historical-evidentialist account suggests that your credence in *Blue* should increase from $1/3$ to near 1 if the treasure is in the blue chest, and to $1/2$ otherwise. I complained that you would thereby violate the norms of dynamic coherence. In response, a defender of the historical-evidentialist approach might seek refuge in an architectural understanding of epistemic norms. On this understanding, you *can't* conform to the norms of ideal dynamic coherence in *Three Chests*.

I have two comments. First, I think friends of the historical-evidentialist account should be reluctant to buy into the architectural conception. The very point of the view is to make the norms distinguish between genuine memory and quasi-memory, and it's not clear if this is an architectural difference. In *Three Chests*, for example, the historical-evidentialist doctrine, too, yields a verdict that is unsatisfiable on the architectural conception: it also demands that your Wednesday

---

19. See [Schwarz, 2025] for the longer story.

credence in *Blue* should depend on what you saw on Monday and Tuesday.

This leads to my second complaint. On the architectural conception, *Three Chests* is a case where you can conform neither to the norms of dynamic coherence nor to the historical-evidentialist norm. Both types of norm describe an unattainable ideal. My complaint is that the "ideal" described by the historical-evidentialist norm is not ideal at all. Ideally, you should retain what you learned; you shouldn't be vulnerable to simple Dutch books, foreseeably reduce your accuracy, or engage in intentionally biased inquiry.

But we don't need to fuss over this. There are yet other kinds of cases in which the evidentialist doctrine clashes with the norms of dynamic coherence, even on a historical conception of evidence, and where compliance with these norms is uncontroversially possible.

7.   Fission Gamble

My next case involves a "fission device" that can turn a human body into several atom-for-atom duplicates, dividing the original body's matter among the duplicates.

> **Fission Gamble**. I have a fission device, 100 identical looking rooms, and a coin that is biased 99/100 towards heads. You and your friend Pedro have agreed to take part in the following experiment.
> I'm about to put you and Pedro to sleep. Then I will toss my coin. If the coin lands heads, I will move you into a randomly chosen one of the 100 rooms, and I will send Pedro into the fission machine, turning him into 99 Pedro copies. Each copy will be moved into one of the remaining rooms. If the coin lands tails, your roles are reversed: I will move Pedro into a randomly chosen room and put 99 copies of you into the other rooms. Once everyone has been moved into their room they are woken up.
> You know all this at the start of the experiment, before you are put to sleep.

Feel free, if you want, to assume that the coin in fact lands heads, so that we don't have to worry about whether the person waking up is the same "you" that I put to sleep at the start of the experiment. My question is how confident you should be, after awakening, that my coin has landed heads.

The conservative answer is simple. Before you were put to sleep, you should have aligned your beliefs with the known chances: your credence in *Heads* should have been 0.99. During the course of the experiment, you don't learn anything that would shed light on the outcome of the coin toss, as judged by your earlier beliefs. If you knew in advance what the 100 rooms look like, you may not receive any unexpected information at all. By the *Expected News is No News* principle, you should retain your confidence in *Heads*.

One might object that people don't survive episodes of fission, which implies that you *do* receive unexpected news upon waking up: that you still exist. If you knew that you will still exist iff the coin lands heads, and now you find that you exist, shouldn't you become certain of *Heads*?[20]

If this were right, it would have striking implications. Physicists and philosophers of physics have argued that we should take seriously an interpretation of quantum mechanics according to which we constantly undergo a kind of personal fission (see, e.g. [Wallace, 2012]). If the objection were correct, dynamic rationality would require being certain that these interpretations are false, no matter their empirical and theoretical virtues.

But the objection isn't correct. It goes wrong by tying the norms of dynamic rationality to the metaphysics of personal identity. [Parfit, 1984] argued that personal identity is not "what matters" from a moral

---

20. Remember that I use second-person locutions like 'you exist' to express self-locating propositions (see note 3). The objection does not assume that your new evidence somehow reveals who you are, which allows you to rule out the fission hypothesis. Rather, the idea is that your old credence was divided between *Heads* possibilities that have a "doxastic successor" and *Tails* possibilities that do not, and that the right way of updating these credences would move all the probability to *Heads* successors.

perspective. The same is true from an epistemic perspective. Even if persons don't survive fission, we may ask how pre-fission beliefs should be related to post-fission beliefs. Whenever a new belief state results from an earlier belief state by an updating process, we may ask how that process should go. There are difficult questions about where to draw the line, but I don't think the answer should be hostage to metaphysical controversies about personal identity. Epistemically, your fission products qualify as "you".[21]

Let's see what the evidentialist doctrine says about *Fission Gamble*. I will divide your evidence, as you wake up, into two parts: *Setup* and *¬Pedro*. *Setup* comprise all the uncentred information you have about the scenario, along with the self-locating information that you have woken up in one of the 100 rooms. I assume that your evidence also entails that you are neither Pedro nor a copy of Pedro: you don't have his (apparent) memories, his glasses, etc. Let *Pedro* be the self-locating proposition that you are either Pedro or a copy of Pedro – for short: that you are "a Pedro". You may have other evidence besides *Setup* and *¬Pedro*, but arguably none of it sheds light on the outcome of my coin toss.

Setup entails that the 100 rooms each contain a person who has just woken up, no matter if the coin has landed heads or tails. Each of the 100 locations is compatible with *Setup*. By a plausible instance of *Self-Locating Indifference*, they have the same evidential probability,

conditional on either outcome of the coin toss.[22] *Setup* also entails that if the coin has landed heads then 99 of the rooms contain a Pedro. So we have $\Pr(Pedro / Heads \wedge Setup) = 0.99$. By the same reasoning, $\Pr(Pedro / \neg Heads \wedge Setup) = 0.01$. Finally, *Setup* entails that the chance of heads is 0.99. So we should have $\Pr(Heads / Setup) = 0.99$. By Bayes' Theorem, it follows that $\Pr(Heads / Setup \wedge \neg Pedro) = 0.5$.

If you proportion your beliefs to your evidence, your credence in *Heads* predictably decreases from 0.99 to 0.5. You violate the principle of conservatism. You also violate other norms of dynamic coherence. You change your beliefs in a way that reduces expected accuracy; you appear to be vulnerable to a simple Dutch Book;[23] you could engage in intentionally biased inquiry.

Distinguishing between genuine memory and quasi-memory doesn't help. To be sure, someone might hold that your evidence includes genuine memories of being put to sleep, and that you could not have such memories if you had been selected to enter the fission device, perhaps because you wouldn't have survived the fission process and genuine memory is tied to the metaphysics of personal identity: you can't have genuine memories of a time when you didn't exist. On this view, your evidence would actually entail *Heads*. But you should not become certain that the coin has landed heads. Your credence in *Heads* should be 0.99, not 1.

---

21. It is important to my arguments in sections 2–4 that your brain-in-a-vat duplicate doesn't qualify as "you". This is admittedly not obvious. We could add some details to the scenario to make it more plausible. To begin, note that it doesn't matter to the earlier discussion if you and the brain differ in some respects, as long as the differences don't affect your evidence. (What this means depends on the concept of evidence that figures in the evidentialist doctrine.) Next, suppose the brain is constructed by first creating a large number of copies of somebody else's brain (say, Pedro's). On Friday, each of these brains is subjected to random changes until one of them is found to be evidentially identical to yours. Then all the other brains are destroyed and a light flashes, signalling that the construction is complete.

22. There is some debate about how information about chance interacts with *Self-Locating Indifference* – see, for example, [Isaacs et al., 2022]. All the rules discussed in that work agree with my assumptions that $\Pr(R=i / Heads \wedge Setup) = \Pr(R=i / \neg Heads \wedge Setup) = 0.01$, where $R=i$ is the self-locating proposition that you are the person in room i (for $1 \le i \le 100$). They also agree with an assumption I'm about to make, that $\Pr(Heads/Setup) = 0.99$. One could avoid the clash between the evidentialist doctrine and the norms of dynamic coherence by assuming that Pr is biased against fission products in such a way that if some evidence is compatible with $N+1$ locations within a world, N of which are occupied by products of fission, then each of these locations has evidential probability $1/2N$ (conditional on the evidence), while the non-fission location has probability $1/2$. But what could motivate this assumption?

23. I say "appear" because it's not entirely clear how we should count your profits or losses in case of *Tails*.

Note that *Fission Gamble* is not a case of evidence loss. Your evidence *deteriorates* insofar as your later evidence is likely to confer a lower probability to the truth than your earlier evidence. But this is not because you lose any evidence. We can assume that whatever was part of your evidence when you were put to sleep remains part of your evidence when you wake up.[24]

Like in *Vat* (and unlike in *Three Chests*), compliance with the norms of dynamic coherence is easily possible in *Fission Gamble*, even on an architectural interpretation. You can preserve your high credence in *Heads*. I say you should. Even though your evidence is neutral on the outcome of the coin toss, you should wake up confident that the coin has landed heads.

8. In defense of diachronic rationality

Ideal rationality, I claim, does not require proportioning one's beliefs to the evidence. Sometimes it requires the opposite. In cases of deteriorating evidence, a rational agent might have to defy their evidence. To illustrate this point, I have described cases where following the evidentialist doctrine would render you "dynamically incoherent": you would violate an attractive principle of conservatism; your beliefs would foreseeably move away from the truth (at least in expectation); you would be vulnerable to simple Dutch Books; you could engage in an irrational activity of manipulating your evidence so as to influence your beliefs.

On reflection, it may not come as a surprise that the evidentialist doctrine can clash with norms of dynamic rationality. Dynamic norms make what an agent should believe at one time depend on their beliefs at other times. The evidentialist doctrine, by contrast, is purely synchronic. If epistemic rationality is simply a matter of proportioning one's beliefs to the present evidence, one's earlier beliefs don't impose

any direct constraints on the later beliefs.

Some have welcomed this implication. Brian Hedden, for example, argues that "being rational is a matter of believing and behaving sensibly, given your perspective on the world", where that perspective is "constituted by your present mental state". [Hedden, 2015, pp.452f.] Past mental states, on this view, can't be directly relevant to what you should do or believe now.

It's easy to see the appeal of this view. Suppose you find yourself at the later time in one of the scenarios I have discussed. What should you believe? What credence function should you adopt? Dynamic norms appear to say that what credence function you should adopt depends on your earlier beliefs. The rule of conditionalization, for example, appears to say that you should take your previous credence function and conditionalize it on your new evidence.[25] But what if you don't have access to your previous credence function? Even if you do, why should your previous credence function play a special role? Suppose you happen to know that Cr_1 is your previous credence function, while Cr_2 and Cr_3 are the previous credence functions of two other people who you regard as epistemic peers or superiors. Why should you adopt an updated version of Cr_1, rather than Cr_2 or Cr_3 or a mixture of the three? (Compare [Christensen, 2000].)

Worries like these motivate what Sarah Moss [2015] calls "time-slice epistemology" – the view that whether an agent is epistemically rational at a time is fully determined by their mental state at that time.

But the worries rest on a misunderstanding. Conditionalization does not say that when you find yourself in a given scenario, you should "take" your previous credence function and "adopt" a revised version of that function. Conditionalization is a rule for how your belief state *at the earlier time* should change as time passes and new information comes in. More generally, dynamic norms put constraints on the dynamics of

---

24. If your earlier evidence includes temporally centred propositions, such as the proposition that you have not yet been put to sleep, then we obviously can't assume that these are still part of your later evidence, which also entails the negation of that proposition. Rather, we might assume that your new evidence entails that you have not been put to sleep *at the earlier time*.

25. I've largely neglected this rule in the previous sections because my cases all involve self-locating information, which conditionalization does not handle well.

an agent's attitudes. They are norms on a process, not on an imaginary choice ("what should I believe?") the agent is assumed to face at the later time. The previous belief state is relevant because it is the starting point of the process. The norms say how *this state* should change. Demanding that a certain belief state should evolve into another does not assume or presuppose that the agent at the later time has access to the earlier state.

None of this requires an especially externalist conception of epistemic rationality. We may well agree that the demands of rationality never turn on contingent facts external to the agent's mind. But minds are temporally extended. They change over time. Rationality can impose constraints on these changes, without thereby turning on external matters.

Time-slice epistemologists reject all (non-trivial) constraints of this kind. In principle, they hold, any mental state may be followed by any other, as long as the states are internally coherent. This is a strikingly revisionary position.

Take a diverse collection of rational agents, in different circumstances, with different beliefs, desires, memories, values, and plans. Now imagine a creature composed by stitching together one-second temporal segments of these agents, in random order. We would not recognize such a creature as an intentional agent, let alone as a rational agent. Most of the psychological properties that make agents candidates of psychological and epistemic evaluation take time. Our everyday conception of a rational agent is a conception not of a time-slice but of a temporally extended agent. We expect that an agent's perceptual experiences influence their subsequent beliefs. We expect agents to utter sentences that don't change grammatical structure and topic or end abruptly after any word. We expect them to engage in extended activities such as eating a meal, going to the library, or reflecting on a difficult choice. We expect them to pursue long-term projects and goals. It is hard to see how these dynamic constraints could be derived from

purely synchronic norms, or why they should.[26]

Even friends of the evidentialist doctrine should be reluctant to subscribe to time-slice epistemology. Consider the following commonplace scenario.

> **Apples.** I put two apples into a previously empty cardboard box. A moment later, I remove one apple from the box. You see all this. You have no reason to think that I'm playing a trick on you. How many apples should you think are left in the box?

Answer: one. You should preserve the information that I initially put two apples into the box, and you should update this information by the new information that one apple has been removed. Without assumptions about how your later mental state should be related to your earlier state, the question could not be answered.

Friends of time-slice epistemology might want to turn the table. They might insist that we really can't say what you should believe in *Apples*. It depends, they might say, on whether you remember that I put two apples into the box, and this is a question on which the norms of rationality fall silent. "Forgetting is not irrational", as [Williamson, 2000, p.219] proclaimed. Dynamic norms like conditionalization, or even the simple principle of conservatism to which I have appealed, are incompatible with forgetting. So much the worse, one might say, for these norms.

I disagree. An ideally rational agent who faces the *Apples* scenario would believe that there is one apple left in the box. If you believe that there are five apples in the box, or zero, or if you profess ignorance, something has gone wrong.

Ideally rational agents don't lose information. Forgetting *is* irrational, in the same way that working on a mathematical problem is irrational. Ideally rational agents already know the answer to all (solvable) mathematical problems, and they never forget.

We, of course, are not ideal. We can't help but forget. Given our

---

26. See [Carr, 2015] for related concerns about time-slice epistemology.

limited cognitive resources, it may even be advantageous to occasionally prune our information, to store only what can be expected to be relevant in the future. For creatures like us, forgetting is rational.

How can one and the same thing be both rational and irrational, both required and forbidden? Epistemic norms, like other norms, are sensitive to feasibility constraints. What is optimal or reasonable for an agent with unlimited cognitive resources may not be optimal or reasonable for creatures like us. The classical norms of Bayesian epistemology ignore cognitive limitations. They require perfect memory, perfect sensitivity to the evidence, and logical omniscience. This ideal is unattainable for creatures like us. If we want to know what *we* should believe, given our cognitive limitations, these norms may not give the right answer. The right answer may be complicated, in part because it depends on empirical details about our limitations.

In this paper, I have focussed on ideal norms. I have argued that ideally rational agents would not lose information, even if their evidence deteriorates. For non-ideal agents, the case against the evidentialist doctrine is even stronger. It is likely that our evidence settles, for example, the answer to Goldbach's conjecture. Yet we should not be certain of the answer.

A wise man does not always proportion his belief to the evidence. Nor should we.[27]

References

Jonathan E. Adler. *Belief's Own Ethics*. MIT Press, Cambridge (MA), 2002.

G.E.M. Anscombe. *Intention*. Harvard University Press, Ithaca, NY, 1957.

Frank Arntzenius. Some Problems for Conditionalization and Reflection. *Journal of Philosophy*, 100(7):356–370, 2003. doi: 10.5840/jphil2003100729.

Frank Arntzenius and Cian Dorr. Self-locating Priors and Cosmological Measures. In Khalil Chamcham, John Barrow, Simon Saunders, and Joe Silk, editors, *The Philosophy of Cosmology*, pages 396–428. Cambridge University Press, Cambridge, 2017.

Nick Bostrom. *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. Routledge, New York, 2002.

Aaron Bronfman. Conditionalization and not Knowing that One Knows. *Erkenntnis*, 79(4):871–892, 2014. doi: 10.1007/s10670-013-9570-0.

Jennifer Rose Carr. Don't Stop Believing. *Canadian Journal of Philosophy*, 45(5-6):744–766, 2015. doi: 10.1080/00455091.2015.1123454.

Jennifer Rose Carr. Epistemic Utility Theory and the Aim of Belief. *Philosophy and Phenomenological Research*, 95(3):511–534, 2017. doi: 10.1111/phpr.12436.

David Christensen. Diachronic coherence versus epistemic impartiality. *The Philosophical Review*, 109(3):349–371, 2000. doi: 10.1215/00318108-109-3-349.

Stewart Cohen. How to be a Fallibilist. *Philosophical Perspectives*, 2: 91–123, 1988. doi: 10.2307/2214070.

Juan Comesaña. Evidentialist Reliabilism. *Noûs*, 44(4):571–600, 2010. doi: 10.1111/j.1468-0068.2010.00748.x.

Earl Conee and Richard Feldman. *Evidentialism*. Clarendon Press, 2004.

Nilanjan Das. Externalism and exploitability. *Philosophy and Phenomenological Research*, 104(1):101–128, 2022. doi: 10.1111/phpr.12742.

Adam Elga. Defeating Dr. Evil with Self-locating Belief. *Philosophy and Phenomenological Research*, 69:383–396, 2004. doi: 10.1111/j.1933-1592.2004.tb00400.x.

Adam Elga. How to Disagree about How to Disagree. In Ted Warfield and Richard Feldman, editors, *Disagreement*, pages 175–186. Oxford University Press, 2007.

J. Dmitri Gallow. Updating for Externalists. *Noûs*, 55(3):487–516, 2021. doi: 10.1111/nous.12307.

Alvin Goldman. What is Justified Belief? In George Sotiros Pappas, editor, *Justification and Knowledge: New Studies in Epistemology*, Philosophical Studies Series in Philosophy, pages 1–23. Springer Netherlands, Dordrecht, 1979. doi: 10.1007/978-94-009-9493-5_1.

Alvin Goldman. Williamson on Knowledge and Evidence. In Duncan Pritchard and Patrick Greenough, editors, *Williamson on Knowledge*, pages 73–91. Oxford University Press, Oxford, 2009.

Alvin Goldman. Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism's Troubles, Reliabilism's Rescue Package. In Trent Dougherty, editor, *Evidentialism and Its Discontents*, pages 254–280. Oxford University Press, Oxford, 2011.

Hilary Greaves. Cluelessness. *Proceedings of the Aristotelian Society*, 116 (3):311–339, 2016. doi: 10.1093/arisoc/aow018.

John Greco. Evidentialism about Knowledge. In Trent Dougherty, editor, *Evidentialism and Its Discontents*, pages 167–178. Oxford University Press, 2011.

Johan E. Gustafsson. *Money-Pump Arguments*. Cambridge University Press, Cambridge, 2022.

Gilbert Harman. *Change in View*. MIT Press, Cambridge (MA), 1986.

Brian Hedden. Time-Slice Rationality. *Mind*, 124(494):449–491, 2015. doi: 10.1093/mind/fzu181.

Sophie Horowitz. Epistemic Akrasia. *Noûs*, 48(4):718–744, 2014. doi: 10.1111/nous.12026.

Colin Howson and Peter Urbach. *Scientific Reasoning*. Open Court Press, La Salle (Ill.), 2nd edition, 1993.

David Hume. *An Enquiry Concerning Human Understanding*. Hackett Publishing Company, Indianapolis, 1777/1993.

Yoaav Isaacs and Jeffrey Sanford Russell. Updating without Evidence. *Noûs*, 57(3):576–599, 2023. doi: 10.1111/nous.12426.

Yoaav Isaacs, John Hawthorne, and Jeffrey Sanford Russell. Multiple Universes and Self-locating Evidence. *The Philosophical Review*, 131(3): 241–294, 2022. doi: 10.1215/00318108-9743809.

Thomas Kelly. Historical versus Current Time Slice Theories in Epistemology. In Brian P. McLaughlin and Hilary Kornblith, editors,

Goldman and His Critics, pages 43–68. John Wiley & Sons, Hoboken (NJ), 2016. doi: 10.1002/9781118609378.ch3.

Matthew Kopec and Michael G. Titelbaum. The Uniqueness Thesis. *Philosophy Compass*, 11(4):189–200, 2016. doi: 10.1111/phc3.12318.

David Lewis. Attitudes *De Dicto* and *De Se*. *The Philosophical Review*, 88 (4):513–543, 1979. doi: 10.2307/2184843.

David Lewis. Elusive Knowledge. *Australasian Journal of Philosophy*, 74: 549–567, 1996. doi: 10.1080/00048409612347521.

Patrick Maher. Probability Captures the Logic of Scientific Confirmation. In Christopher Hitchcock, editor, *Contemporary Debates in Philosophy of Science*, pages 69–93. Blackwell, Oxford, 2004.

Kevin McCain. The Virtues of Epistemic Conservatism. *Synthese*, 164 (2):185–200, 2008. doi: 10.1007/s11229-007-9222-5.

Kevin McCain. Is Forgotten Evidence a Problem for Evidentialism? *The Southern Journal of Philosophy*, 53(4):471–480, 2015. doi: 10.1111/sjp. 12152.

John McDowell. *Perception as a Capacity for Knowledge*. Marquette University Press, Milwaukee, 2011.

Sarah Moss. Time-slice Epistemology and Action under Indeterminacy. *Oxford Studies in Epistemology*, 5:172–94, 2015. doi: 10.1093/acprof: oso/9780198722762.003.0006.

Derek Parfit. *Reasons and Persons*. Clarendon Press, Oxford, 1984.

Richard Pettigrew. *Accuracy and the Laws of Credence*. Oxford University Press, 2016.

Richard Pettigrew. *Epistemic Risk and the Demands of Rationality*. Oxford University Press, Oxford, 2022.

Bernhard Salow. The Externalist's Guide to Fishing for Compliments. *Mind*, 127(507):691–728, 2018. doi: 10.1093/mind/fzw029.

Jonathan Schaffer. Contrastive Knowledge. *Oxford Studies in Epistemology*, 1:235–271, 2005. doi: 10.1093/oso/9780199285891.003.0009.

Miriam Schoenfield. Conditionalization Does Not (in General) Maximize Expected Accuracy. *Mind*, 126(504):1155–1187, 2017. doi: 10.1093/mind/fzw027.

Wolfgang Schwarz. Sleeping Beauty and the Demands of Non-Ideal

Rationality. *Noûs*, Forthcoming, 2025.

Brian Skyrms. A Mistake in Dynamic Coherence Arguments? *Philosophy of Science*, 60(2):320–328, 1993. doi: 10.1086/289735.

Martin Smith. *Between Probability and Certainty: What Justifies Belief*. Oxford University Press, 2017.

Katie Steele and H. Orri Stefánsson. Belief Revision for Growing Awareness. *Mind*, 130(520):1207–1232, 2021. doi: 10.1093/mind/fzaa056.

Hamid Vahid. Varieties of Epistemic Conservatism. *Synthese*, 141(1): 97–122, 2004. doi: 10.1023/B:SYNT.0000035849.62840.e8.

David Wallace. *The Emergent Multiverse: Quantum Theory According to the Everett Interpretation*. Oxford University Press, Oxford, 2012.

Brian Weatherson. Memory, Belief and Time. *Canadian Journal of Philosophy*, 45(5-6):692–715, 2015. doi: 10.1080/00455091.2015.1125250.

Timothy Williamson. *Knowledge and Its Limits*. Oxford University Press, Oxford, 2000.