Don't Stop Believing

Epistemic rationality requires two kinds of coherence. Broadly speaking, an agent's beliefs must fit well together at a time, and also fit well together over time. At any particular time, we should avoid believing contradictions, believe the consequences of our beliefs, and so on. And over time, we should respect the evidence we've received and adapt our beliefs to new evidence.

The traditional Bayesian picture of epistemic rationality is simply the conjunction of a synchronic claim and a diachronic claim:

Synchronic coherence: Rational belief states form a probability function and are rationalized by one's evidence.

Diachronic coherence: Rational belief states evolve by retaining old certainties and conditioning on new evidence.

Recently, however, a number of philosophers have pushed for the abandonment of diachronic norms. Norms like Conditionalization, that have historically been understood as constraints on beliefs at different times, have been reinterpreted as purely synchronic constraints. According to this view, the norms of rationality, practical or epistemic, apply only to time-slices of individuals.

I want to resist this movement. I'll argue for the following claim:

Diachronic Rationality: There are diachronic norms of epistemic rationality.

The problem that the opponent of diachronic rationality poses is this: diachronic norms of epistemic rationality are in tension with *epistemic internalism.* Epistemic internalism, in its most generic form, is the view that whether or not you're epistemically rational supervenes on facts that are 'internal' to you. The relevant sense of 'internal' can be cashed out in a variety of ways. If there are diachronic norms of epistemic rationality, then whether you're epistemically rational *now* is determined in part by your past epistemic states. And facts about the past are not, in the relevant sense, internal to you.

The proponent of diachronic norms faces a dilemma. We can't endorse both of the following claims: that epistemic rationality imposes cross-temporal constraints

on belief, and that epistemic rationality is determined only by what is 'internal' to the agent.

Faced with a choice between diachronic norms and epistemic internalism, I will argue that we should choose diachronic norms. I argue that that the rejection of diachronic norms incurs a number of serious problems: most notably, that it permits discarding evidence, and that it treats agents who are intuitively irrational as epistemic ideals.

Here is how the paper will proceed: in section 1, I'll explain the framework in which much of my discussion takes place, i.e., the Bayesian view of rationality. Then I'll introduce in more detail the objection to diachronic epistemic norms, some of its common motivations, and how the debate is situated within epistemology.

In section 2, I offer three objections to the synchronic-norms-only view. In 2.1, I argue that time-slice rationality entails that discarding evidence is rational. 2.2 argues that there are intuitive normative differences between agents who conform to diachronic norms and those who don't. The opponent of diachronic norms is committed to a strong claim: that no agent can ever be worse than another in virtue of purely diachronic differences between them. There are intuitive counterexamples to this generalization. In 2.3, I argue that according to an attractive view in philosophy of mind, all irrationality is fundamentally diachronic. So the synchronic-norms-only view may wind up committed to there being no epistemic rationality at all.

In section 3 I discuss the motivates, explicit or tacit, of the synchronic-norms-only view. I discuss the idea that cognitive limitations somehow limit our epistemic liability in 3.1. In 3.2 I discuss the idea of epistemic *ought*-implies-*can* and epistemic *responsible*-implies-*can*. 3.3 describes a notion of relative rationality, which allows us to accommodate many of the intuitions cited in favor of the synchronic-norms-only view.

Section 4 discusses an objection to diachronic norms prohibiting information loss. What if one can ensure a net gain in information only at the cost of losing some information? I discuss diachronic norms that can accommodate the idea that this sort of 'information trade-off' can be rational. I conclude briefly in section 5.

1 The conflict

1.1 Bayesianism

Before I begin, let me state some background assumptions. First, I will assume a partial belief framework. (Nothing hinges on this.) On this view, beliefs come in

degrees (where a degree of belief is called a 'credence'). Credences fall in the interval $[0,1]$, where credence 1 represents certain belief, credence 0 represents certain disbelief, credence $\frac{1}{2}$ represents maximal uncertainty, and so on. A person's total belief state is represented by a credence function, i.e. a function from propositions to real numbers in $[0,1]$.

According to the classical Bayesian picture, there are two kinds of coherence that rational credences exhibit, one synchronic and one diachronic. The synchronic constraint is known as Probabilism:

Probabilism: Rational credences form a probability function: that is, they obey the following three axioms. Where $\mathcal{W}$ is the set of all worlds under consideration[1]:

1. *Nonnegativity:* for all propositions $A \subseteq \mathcal{W}$, $Cr(A) \geq 0$

2. *Normalization:* $Cr(\mathcal{W}) = 1$

3. *Finite additivity:* if A and B are disjoint, then $Cr(A \vee B) = Cr(A) + Cr(B)$

The diachronic constraint is known as Conditionalization:

Conditionalization: let E be the strongest proposition an agent learns between t and t'. Then the agent's credences should update such that $Cr_{t'}(\cdot) = Cr_t(\cdot \mid E)$, where $Cr(A \mid B)$ is usually defined as follows:

$$Cr(A \mid B) = \frac{Cr(A \wedge B)}{Cr(B)}$$

Conditionalization has two basic effects: first, you treat all possibilities (that is, worlds) that are incompatible with your new evidence as dead. They are given credence 0. Second, you reapportion your credences among the remaining live possibilities, preserving relative proportions between the possibilities.

Now, one of the consequences of Conditionalization is that once you rationally learn something, you can't rationally unlearn it. You can't rationally lose information. (The set of live possibilities only shrinks.) This is, as stated, a strong and fairly controversial constraint.

There are analogs to Conditionalization in the full belief framework. For example, Jane Friedman (manuscript), defends the following norm of inquiry: when a question has been closed, don't reopen it. This is a close analog to Conditionalization's controversial consequence: that possibilities with credence 0 cannot recover

---

[1] Throughout I will be assuming that credence functions range over subsets of a finite set of worlds.

positive probability. There are other diachronic norms that are weaker: for example, some forms of epistemic conservatism say that if you rationally believe a proposition at an earlier time, then it remains rational for you to continue believing it at later times, as long as you don't receive any new, disconfirming evidence.

I want to offer a general diachronic norm that cross-cuts whether we treat belief states with the full belief framework or the partial belief framework, and also cross-cuts whether we treat the overriding diachronic norm as Conditionalization, or whether we accept alternatives diachronic norms on credences (e.g. Jeffrey Conditionalization). Here is a candidate:

Diachronic evidentialism: An agent should only change her epistemic state by updating on new evidence.

Note that this is, on its face, a fairly strong norm. One needn't endorse this strong a norm in order to believe that there are diachronic constraints on rationality. But we'll start with something this strong, and see what can be said in favor of it.

First, though, we should consider objections to diachronic norms.

1.2 The rejection of diachronic rationality

Sarah Moss (manuscript) describes a 'general movement' towards rejecting diachronic norms of rationality. The aim of this movement: to take statements of diachronic norms like Conditionalization and replace them with analogous synchronic norms. According to Moss:

> It is naïve to understand Conditionalization as a diachronic rule that says what credences you should have at a later time, given what credences you had at an earlier time, literally speaking. Instead we should understand it as a synchronic rule... Of course, one might claim that Conditionalization was originally intended as a literally diachronic rule, and that 'Conditionalization' should therefore be reserved for a rule that binds together the credences of different temporal slices of agents—but I am inclined to interpret the Founding Fathers charitably. (Moss manuscript, 24)

Opponents of diachronic epistemic norms include Talbott (1991), Christensen (2000), Williamson (2000), Meacham (2010), and Hedden (2012).

There are a variety of motivations for a synchronic-norms-only epistemology. Some, e.g. Williamson, simply find diachronic constraints like Diachronic Evidentialism implausible. For others, the synchronic-norms-only view follows from a more general principle—in particular, some form of epistemic internalism. Here, for example, is Meacham (2010):

> In Bayesian contexts, many people have appealed to implicitly internalist intuitions in order to support judgments about certain kinds of cases. But diachronic constraints on belief like conditionalization are in tension with internalism. Such constraints use the subjects beliefs at other times to place restrictions on what her current beliefs can be. But it seems that a subjects beliefs at other times are external to her current state. (87)[2]

There are a number of different forms of epistemic internalism. The two varieties that are perhaps most familiar are *mentalist internalism* and *access internalism.*

Mentalist Internalism: the facts in virtue of which a subject is epistemically rational or irrational supervene the subject's mental states.[3]

Access Internalism: the facts in virtue of which a subject is epistemically rational or irrational supervene on those of the subject's mental states that she's in a position to know she is in.

It's worth noting that neither of these immediately conflicts with diachronic constraints on rationality, at least as stated. After all, it might be that what's rational for an agent believe at one time supervenes on her mental states at another time, or her mental states at many different times, or those mental states that she has access to at many different times, etc.

Opponents of diachronic norms often appeal to a form of access-internalism: facts about our past mental states are irrelevant to our current rationality because they are, at least in some circumstances, inaccessible to us.[4] (A mental state is *accessible* to an agent iff, if the agent is in the mental state, then she is in a position to know that she is.) And so the internalist objection to diachronic rationality is best interpreted as involving the following form of internalism:

---

[2] Note that while Meacham argues that there is a conflict between Conditionalization and internalism, and provides a synchronic alternative to Conditionalization, he is (at least in his (2010) not committed to the denial of traditional diachronic Conditionalization.

[3] Note that this is (at least arguably) orthogonal to internalism about mental content. It's consistent to hold that whether an agent's beliefs are rational is determined by what's in the head, while at the same time holding that the correct characterization of the *contents* of an agent's beliefs will involve appeal to the agent's environment.

[4] Williamson is, of course, an exception, since he is not an internalist of any sort. Christensen's objection to diachronic norms, which I discuss in section 4, doesn't require appeal to any form of internalism.

Time-Slice Internalism: the facts in virtue of which a subject is epistemically rational or irrational *at a particular time t* supervene on those of the subject's mental states that she's in a position to know she is in *at t*.

Here's an example statement of this sort of internalism:

> Whether it is rational to retain or abandon a belief at a time is a matter of which of these makes sense in light of your current epistemic perspective, i.e., in light of what you currently have to work with in revising your beliefs. (McGrath 2007, 5)

Time-slice internalism immediately entails that the norms governing epistemic rationality are purely synchronic.

The motivations for time-slice internalism draws on an analogy between the past and the external: our access to our past mental states is, at least in principle, limited in just the same way as our access to the external world.[5] The fact that we had certain mental states in the past does not entail that we are, at present, in a position to know that we had those mental states.

We can show the differences between time-slice internalism and traditional access internalism by appeal to different forms of skeptical scenario:

Example #1

Suppose there are two agents who have exactly the same mental states. Furthermore, both agents have access to exactly the same mental states. But one agent has mostly true beliefs about the external world; the other is a brain in a vat and is systematically deceived about the external world.

The internalist intuition about this case: if the undeceived agent is rational, so is the brain in the vat.

The time-slice internalist invites us to make the analogous judgment about an agent who is systematically deceived not about the external world, but about her past memories:

Example #2

Suppose there are two agents who have exactly the same mental states at a particular time *t*. Furthermore, both agents have access to exactly the same mental states. But one agent has mostly true beliefs about her

---

[5] Meacham (2010), Hedden (2012).

past memories; the other has a brain implant that dramatically alters her beliefs, memories (or, if you like, quasi-memories), and other mental states erratically, and so at t she is systematically deceived about her past beliefs.

The question is: should these cases be treated as epistemically analogous? Do we have the same kind of intuition that, in the second example, if the ordinary agent is rational, then the memory-scrambled agent is rational? I would find it surprising if anyone claimed to have strong intuitions about whether the latter agent is rational.

The proponent of synchronic-norms-only rationality emphasizes the analogy between the agent who's deceived about the external world and the agent whose memories are regularly scrambled. After all, they are both doing the best they can under strange, externally imposed circumstances.

The proponent of diachronic norms responds that the scrambled agent should instead be understood on analogy to someone who is given a drug that makes him believe contradictions. They are both doing the best they can under strange, externally imposed circumstances—but nevertheless, they are not ideally rational. I'll argue for this claim in greater detail in section 3.

1.3 Orienting the debate

I'm concerned to defend a fairly weak claim: that there are diachronic norms of epistemic rationality. Advocating diachronic epistemic norms does not entail advocating Conditionalization, which is clearly an extremely strong constraint.

To orient the debate over diachronic norms, we can consider various kinds of loose (!) alliances. The debate is in some ways aligned in spirit with the debate over epistemic externalism v. internalism, for obvious reasons: if there are genuinely diachronic epistemic norms, then whether a belief state is rational at a time can depend on facts that are inaccessible to the agent at that time.

There are also some similarities in spirit between defenders of diachronic norms and defenders of epistemic conservatism. According to epistemic conservatism (at least, of the traditional sort; there are, of course, varieties of conservatism), if you find that you have a belief, that provides some (defeasible) justification for continuing to have that belief. One way of drawing out this analogy: the epistemic conservatist holds that if an agent rationally believes that p at t, then it is (ceteris paribus) *permissible* for the agent to believe that p at a later t'.[6] The defender of a diachronic norm like Conditionalization holds that if an agent rationally believes that p (with certainty) at t, then she is rationally *required* to believe that p at t'.

---

[6] See e.g. (Burge 1997).

7

But it's worth noting that there are weaker diachronic requirements that could constrain rational belief: for example, that one shouldn't reduce or increase confidence in a proposition (in which her previous credence was rational) unless she receives new evidence *or* forgets evidence. The time-slice internalist is, therefore, endorsing a fairly strong claim. As I'll argue in the next section, there are costs to denying that rationality imposes *any* diachronic constraints on belief.

2   Problems for time-slice rationality

2.1   Problem #1: permissibly discarding evidence

One of the benefits that time-slice internalists claim for their view is that, by rejecting Conditionalization, they are able to vindicate the idea that forgetting doesn't make a person irrational. If Conditionalization applies, without qualification, over the whole of an agent's life, then any instance of forgetting would be sufficient to make the agent irrational.

The flip side is that time-slice internalism also makes any instance of discarding evidence epistemically permissible. And discarding evidence is a canonical example of a violation of epistemic norms. The reason that time-slice internalism has this effect is that discarding evidence is a fundamentally diachronic phenomenon. At some time, you receive evidence. At a later time, your attitudes fail to reflect the fact that you've received that evidence.

> **Example #3**
>
> Suppose an agent has strong beliefs about whether capital punishment has a deterrent effect on crime. Then he learns of a study that provides evidence against his view. So he should reduce his confidence in his belief. But instead our agent (involuntarily) discards the evidence; he loses any beliefs about the study; it has no enduring effect on his attitudes regarding capital punishment. Now he can go on confidently endorsing his beliefs without worrying about the countervailing evidence.

This is a standard example of irrationality. (One might object: an agent like this is epistemically irrational only if he voluntarily discards the evidence. But cognitive biases are not voluntary; so this objection would have the consequence that cognitive biases never result in irrational belief. I take this to be uncontroversially false.)

Discarding evidence is epistemically irrational. Therefore there are diachronic norms of epistemic rationality. There's not much more to say about this. But to

my mind it is a serious challenge to the synchronic-norms-only view; perhaps the most serious.

2.2 Problem #2: deviating from epistemic ideals

Some kinds of belief change are plausibly described as deviating from some sort of epistemic ideal, even when no synchronic norms are violated. It might be controversial whether, by virtue of deviating from the ideal, the agent is irrational. But given that there are purely diachronic epistemic ideals to deviate from, it follows that there are diachronic epistemic norms.

Consider again an agent whose total belief state is entirely overhauled at regular, and perhaps frequent, intervals (every minute? every second?). At every instant her credences are probabilistically coherent. And they uphold any other synchronic constraints on rational belief: for example, they are appropriately sensitive to chance information; they reflect whatever the epistemically appropriate response is to whatever phenomenological inputs the agent has at that instant; etc. However strong you make the norms of synchronic rationality, our agent obeys all of those norms at each instant.

But her total belief state at one moment is largely different from her total belief state at the next. If you asked her a minute ago where she was from, she'd say Orlando; if you asked her now, she'd say Paris; if you ask her a minute from now, she'll say Guelph. These changes are random.

The time-slice internalist is committed to the claim that our agent is *ideally rational.* I think this is false. Whether or not the agent rises to the level of rationality, it is clear that she is epistemically sub-ideal: she is doing worse, epistemically, than someone whose credences are more stable over time.[7]

Objection: If her evidence changes with each belief overhaul, then perhaps it is rational for her to overhaul her beliefs so frequently.

Reply: In order to assess whether her evidence changes with each belief overhaul, we would need to say more about what 'her evidence' is. For example, if you believe her evidence is what she knows,[8] then no doubt it'll overhaul, since her beliefs overhaul. That doesn't get us any further toward defending the claim that she is rational. It might just be that she irrationally stops believing various propositions that she previously knew.

---

[7] Of course, it's entirely appropriate that an agent's beliefs should continuously change *a little* all the time: she should update on new information about, e.g., the passage of time, new evens that she encounters, etc. But in the example I'm concerned with, a much greater proportion of her beliefs change, and not simply because she's exposed to new evidence.

[8] See (Williamson 2000) for the canonical defense of this identity.

If you believe her evidence is something else—perhaps something to do with phenomenology—then in order to ensure that the example is one where she obeys synchronic norms, this will have to overhaul regularly too. But let's note that phenomenology a very, very thin source of information. If you think that memorial phenomenology is the basis for your beliefs—consider how few of those beliefs can be justified on the basis of memorial phenomenology at any given moment. There is a limit to how much phenomenology you can conjure up in a short time span. To the extent that I understand this sort of view, it seems to me that it is susceptible to the same charge that the time-slice internalist presses against the defender of Conditionalization: that it declares us all irrational.

On the other hand, if diachronic evidentialism is correct, then 'her evidence' is all the evidence she has received, not just the evidence that is accessible to her in the moment. The diachronic will say: her evidence does not dramatically change, and therefore it's irrational for her beliefs to dramatically change.

Now, one can agree with me that the agent with erratically shifting beliefs is epistemically non-ideal, and still judge the agent to be rational. One might, for example, have a *satisficing* view of rationality: maybe it isn't necessary to perfectly satisfy all epistemic norms in order to be epistemically rational. This kind of view isn't common among Bayesians, who tend to accept that rationality just is ideal rationality, and who tend to accept happily that none of us is rational. But I take it that this is a common presupposition in informal epistemology. For example, informal epistemologists typically accept that it's not rationally required that we believe all the consequences of our beliefs, though we would be rationally better if we did.

As long as we accept that the agent whose credences shift erratically is doing worse, epistemically speaking, than the agent whose credences only change by rational update on evidence, then for my purposes, it doesn't matter whether we call the former agent irrational or rational. We have already admitted that there are diachronic epistemic norms.

2.3 Problem #3: all incoherence is diachronic

Some form of functionalism about belief is widely held among formal epistemologists. Probably the most influential form is **interpretivism**, the view that your belief state is simply the body of credences that best rationalize your behavioral dispositions.

One effect of this sort of functionalism is that the kinds of facts in virtue of which you believe A at a particular time t are also facts in virtue of which you don't believe $\neg A$ at t. Similarly, the facts in virtue of which you have credence .6 in B at t also make it the case that you have credence .4 in $\neg B$ at t. The facts that

make it the case that you've received evidence that *C* at *t* are facts that make it also the case that you increase your credence in *C* at *t*. And so on.

How, then, can we ever correctly attribute incoherence to anyone? Agents can be 'fragmented': they can, in effect, believe different things relative to different contexts, or for different purposes. For example, an agent may exhibit a belief that *A* in her linguistic behavior in some contexts, and yet manifest a belief that ¬*A* in her non-linguistic actions in another context.

One needn't accept the full fragmentationist package in order to accept that the very facts that make it the case that a person believes *A* at a particular time also make it the case that he rejects ¬*A* at that time; and similarly for other sorts of relations that beliefs stand in.

An effect of this view: in a particular context, at a particular time, an agent is *always* synchronically coherent. Synchronic coherence, on this interpretation, is either a trivial norm, or else a norm constraining belief-attribution rather than belief itself.

If this view in philosophy of mind is correct, opponents of diachronic rationality are pushed in to a corner. They must either reject this attractive philosophy of mind, or else reject the idea that there are *any* substantive epistemic constraints on belief. Neither of these is an attractive option.[9]

3   Epistemic 'blamelessness' does not entail epistemic ideality

3.1   Diachronic evidentialism and information loss

Let me emphasize again: I am concerned primarily with defending Diachronic Rationality, the claim that there are diachronic epistemic norms. Here are two stronger claims:

Rationality = Ideal Rationality In order to be epistemically rational, one must perfectly satisfy all epistemic norms, synchronic or diachronic;

or even stronger:

---

[9] *Objection:* fragmentationists might also hold that individual fragments also necessarily update by Conditionalization, and so for each fragment this norm is also trivial. *Reply:* One can exhibit irrationality diachronically in other ways than by failing to conditionalize *within* a fragment. For example, fragments can conflict with each other in such a way as to make the agent irrational. But the facts in virtue of which this is true have to do with different fragments characterizing the agent's belief state manifest at different times.

Rationality Requires Lifelong Conditionalization In order to be epistemically rational, you must satisfy Conditionalization over the entire course of your life.

There are a variety of ways we could resist these extensions of Diachronic Rationality. For example, you might accept that satisfying Conditionalization would make an agent epistemically better but that it isn't always necessary for rationality; perhaps there are sometimes extenuating circumstances. Or you might accept that Conditionalization is rationally required over stretches of time, but not an agent's entire life. (Perhaps it's required between instances of some psychological event of forgetting, where this might be psychologically distinguished from discarding evidence.)

The real question is: who is making the universal claim and who is making the existential claim? The opponent of diachronic norms insists that no one is *ever* irrational by virtue of diachronic facts. All we need to convince ourselves that this view is false is one instance where, e.g., discarding evidence is epistemically sub-ideal.

Nevertheless, I want to explore a defense of diachronic evidentialism, the comparatively strong claim that epistemically, we should only change our beliefs by updating on new evidence. We'll set aside the question of whether an agent who violates diachronic evidentialism is irrational in all circumstances.

One form of diachronic evidentialism is Conditionalization. A common complaint against Conditionalization is that it entails that forgetting something learned with certainty is irrational. This result is often met with an incredulous stare; counterargument is treated as unnecessary.

> Forgetting is not irrational; it is just unfortunate. (Williamson 2000, 219).

It seems to me that forgetting is not just unfortunate but *epistemically* unfortunate. And 'epistemic misfortune' is simply a gentler name for epistemic sub-ideality. In any case, even if Williamson is correct, it may still be that Conditionalization has epistemic normative force.[10]

Let me acknowledge: I'm not concerned about whether we accept the claim I above called 'Rationality = Ideal Rationality.' Where we draw the line between

---

[10] There are complaints against Conditionalization that have nothing to do with forgetting: for example, that it only allows update when our evidence provides us with credence 1 in some new proposition (unlike, e.g., Jeffrey Conditionalization), and that it doesn't allow us to lower our credence in any proposition from 1 even in circumstances where no forgetting takes place (e.g. in Artnzenius's (2003) Shangri-La example). But neither of these objections extends to Diachronic Evidentialism; so these considerations simply invite us to find a suitable diachronic replacement for Conditionalization.

epistemic trespasses that are is sufficient for irrationality and those that aren't doesn't seem to me obviously substantive. Sociologically speaking, formal and informal epistemologists tend to talk about rationality in quite different ways. For many informal epistemologists, to be 'irrational' is to be (at least a little) insane; the majority of us are by and large rational. It is common to think, e.g., that one is not rationally required to believe all the consequences of one's beliefs (even though perhaps by doing so you'd be epistemically better). By contrast, among formal epistemologists, it is more common to use 'irrational' to mean *rationally imperfect*. To be epistemically 'irrational', in their sense, is to deviate from epistemic ideals.

Now, whether or not we call it 'irrational', forgetting—losing information—deviates from our epistemic ideals. Compare it with other epistemic ideals:

Deductive closure: if an agent believes A and $A \vdash B$, then the agent should believe B.

Because of our cognitive limitations—for example, the fact that we can't believe all mathematical truths—actual agents' beliefs are never actually closed under deduction.

Probabilism creates similar problems:

Probabilism. Our credences must form a probability function.

This entails that we must, e.g., have credence 1 in all necessary truths. It also entails that we must have infinitely precise credences: that there be a difference between having credence .2 and credence .20000000000001. But because of our cognitive limitations (we are finite beings!), actual agents never actually have infinitely precise credences.

It should be a familiar point, then, that because of our cognitive limitations, no actual agents are epistemically ideal. And there's no obvious reason to treat forgetting any differently. Actual agents' forgetfulness is just another cognitive limitation that stands in the way of epistemic ideality.

3.2 Epistemic *ought* implies *can*?

Now, one might object: so much the worse for any of these norms! Surely we're not blameworthy for beliefs that result from our cognitive limitations. If you *can't* satisfy the norm, then the norm doesn't apply to you. (After all, *ought* implies *can*.)

But this is simply false. Our friend in his tinfoil hat can't make himself stop overtly believing contradictions. That doesn't make him *epistemically ideal.* It is

a commonplace in epistemology that sometimes a person can be irrational even when he is 'doing the best he can'.

Even if the epistemic *ought*-implies-*can* argument were successful against ideals like deductive closure, probabilism, or precise credences, it's not clear how it is supposed to apply to forgetting. After all, a norm against forgetting would say: if you're in a certain kind of state, you should *continue* to be in that state. In the other cases, no actual agents can be in the relevant state in the first place. So it's not as though it's psychologically or physically impossible for you to be in the recommended belief state. It's just that you can't always *make yourself* remember something.

But in epistemology *ought* ϕ doesn't imply *can make yourself* ϕ. It's not as though you can simply make yourself believe anything. (Try believing that I am a goat!) Beliefs are not under our immediate voluntary control.[11] And so if there's any sense in which *ought*-implies-*can* in epistemology—which is doubtful—it does not apply in the case of forgetting.

This point generalizes to any argument for time-slice internalism that appeals to the idea that we cannot be responsible or blameworthy for believing in accordance with past evidence that we no longer have immediate access to. Epistemic rationality has nothing to do with responsibility or blameworthiness.

Perhaps the greatest challenge for the time-slice internalist is to justify their view in some way that doesn't appeal to some misguided epistemic *ought*-implies-*can* or epistemic *ought*-implies-*responsible* principle.[12]

3.3 Relative rationality

One fear we might have about accepting epistemic principles that ordinary agents can't perfectly realize is that we would then have to accept that the norms of rationality are, in some sense, only for ideal agents; they don't apply to any actual agents.

But that's rather like saying that if you're not ideally law-abiding—you've already gotten a speeding ticket; there's nothing you can do to change that fact—then traffic laws no longer apply to you. Suppose the traffic laws say:

1. Don't get speeding tickets;

---

[11] We can take actions to induce beliefs, e.g. gathering evidence, or take actions to slowly indoctrinate ourselves over time. But generally speaking, we cannot believe something merely by trying.

[12] In a 2012 AAP talk (no manuscript currently exists), Wolfgang Schwarz argued, similarly, that the motivtion for rejecting diachronic norms derives from the idea that they cannot be action-guiding, and this turns on an illicit conflation of the practical with the epistemic.

2. If you get speeding tickets, pay the speeding tickets;

3. If you don't pay your speeding tickets, go to your court hearing;

4. . . .

Then this set of legal norms generates different 'levels' of law-abidingness. 'Ideal law-abidingness' amounts to obeying *all* of these (where everything after 1 you satisfy trivially by virtue of satisfying 1). Still, if you *can't* obey all of the laws, you're legally required to obey 2, 3, . . . ; and if you *can't* obey 2, then you're legally required to obey 3, etc.. What the traffic laws require of you in particular circumstances is relativized to what you are capable of. Still, though, if you are not capable of satisfying all of the laws, then you are not *ideally* law-abiding.

We can represent the norms of rationality as having a similar structure:

1. Be diachronically *and* synchronically coherent.

2. If you can't be both, be synchronically coherent.

3. . . . etc.

So, like law-abidingness, we can think of rationality as *relative*—in particular, relative to our cognitive limitations. Ideal rationality is a special case of relative rationality: it is the case where there are no limitations.

3.4 Rationality vs. epistemic ideality?

I have emphasized that there's a clear sense in which the subject who violates diachronic norms is doing worse, epistemically, than the subject who doesn't. But the time-slice internalist might object: the person who happens to *know* less is also doing worse, epistemically, than a person who knows more. But that doesn't mean that the person who knows less is *irrational*. So, the time-slice internalist might conclude, not all epistemic norms are norms of rationality.

There is a natural way of drawing a distinction between norms of epistemic rationality vs. other epistemic norms. In the practical realm we sometimes distinguish 'objective' and 'subjective' norms. By analogy, we might consider it an objective epistemic norm that we should believe all true propositions and disbelieve all false propositions, in the same way that according to the utilitarian, it is an objective norm that we should maximize utility. And conversely, we might think of norms of rationality as including only the subjective norms. Where do diachronic norms fall on this divide? Which of the epistemic norms are norms of epistemic rationality?

As a working hypothesis, here is my suggestion: we should think of the norms of epistemic rationality as those that characterize the states of the *agent* and not her environment. One of the ways of cashing this out: the epistemic norms are the constraints that characterize the epistemic states of the ideal information gatherer.[13] The ideal information gatherer is non-omniscient; none of her beliefs is guaranteed to be true except on the basis of evidence.[14]

Epistemic rationality involves having beliefs that approximate the truth as much as possible, given our non-omniscience. On this view, though, there's no reason to think of diachronic norms as somehow external to rationality. Retaining information will, by and large, help you keep your belief state more accurate.

4 Rational information loss

4.1 Losing information to gain information

Now, it can't be that losing information *necessarily* makes your beliefs less accurate. For example: suppose that, by chance, you happen to forget only misleading evidence. Then losing information actually makes your beliefs more accurate. Rather, retaining information makes it more likely that your credences will be more accurate, roughly speaking. It increases the *expected accuracy* of your credences. (I will say more about this in section 4.2.)

Now, conditionalizing on new information is an example of pure information gain. And forgetting and discarding evidence are examples of pure information loss. But what should we say about mixed cases?

We can define an **information trade-off** as a case where you gain some information at the cost of losing some other information. If taking an information trade-off can be rational, then some forms of diachronic norm are false. For example, Conditionalization is false: rational informational trade-offs would require rational information loss. Christensen (2000) uses an example with the following structure to argue against the view that there are diachronic epistemic norms:

---

[13] Schwarz defended Conditionalization with this analogy: suppose we want to build a robot to gather information for us in whatever environment he ends up in. We have the option of programming it to obey diachronic evidentialism. Should we? It seems fairly obvious that we should: then the robot will not lose information, and so will end up with more information.

[14] This is, again, a working hypothesis. But there is another answer that I'm sympathetic to. There's a few suggested in semantics, but separable from linguistic considerations, that there is no real subjective/objective divide. There are simply grades of more or less subjective norm. Likewise, it might be that there's ultimately no defensible line to be drawn between subjective and objective epistemic norms. Diachronic norms are more 'subjective' than the norm of truth, but more 'objective' than some synchronic norms (like probabilism), which are in turn more 'objective' than other synchronic norms, and so on.

Example #4

Suppose you know that someone knows more than you about some topic. You know some things she doesn't know, but on the whole she's more informed on the topic. It would be gauche to ask her about the topic. Luckily, you have the option of using a credence downloader to replace your credences on the topic with hers. Is it permissible for you to do so?

Christensen invites us to judge that it is indeed permissible.

Now, it should be clear that this is at best an argument against *some* diachronic norms, not against diachronic rationality in general. But one interesting fact about this case is that if you take the trade-off, you violate Conditionalization—but you also increase the expected accuracy of your credences. So, if epistemic rationality consists in maximizing expected accuracy, then Conditionalization can't be a norm of epistemic rationality.

Now, there are two possible objections one could make against Conditionalization on the basis of an example like this.

Objection #1. Taking the trade-off maximizes expected accuracy, so you're rationally required to violate Conditionalization.

This shouldn't trouble the proponent of Conditionalization. The norms of epistemic rationality govern only epistemic states, not actions like using a credence downloader. If we were rationally required to perform actions that maximize the expected accuracy of our credal states, then we would, for example, be rationally required to perform constant experiments, to read all of Wikipedia, etc.

Objection #2. If you do take the trade-off, your resulting epistemic state is rational. So it must be permissible to violate Conditionalization.

This objection is more troubling for the proponent of Conditionalization. If this objection is correct, then Conditionalization is false. At most Conditionalization holds across periods of time where no informational trade-offs are available.

There are the two options, then, for the proponent of diachronic norms:

1. We can stick with Conditionalization and reject the claim that there are epistemically rational informational trade-offs. (We might concede that informational trade-offs are still pragmatically rational.)

2. Alternatively, we can adopt diachronic norms that are more liberal that Conditionalization.

There's something to be said for both options and I won't defend one over the other. There's little more to be said about the first option, though, so let's explore the second option. But first, we should say a little bit more about what expected accuracy is.

4.2   Epistemic utility theory

Epistemic utility theory formalizes the notions of the accuracy and the expected accuracy of a credence function. The aim of epistemic utility theory is to use the tools of decision theory, combined with an epistemic version of value, in order to give a foundational justification for various epistemic norms.

The most widely discussed epistemic utility functions are *gradational accuracy* measures. The accuracy of a credence is its nearness to the truth (by some measure). A credence function with maximal accuracy would assign credence 1 in all truths and credence 0 in all falsehoods. In other words, it would be omniscient.

Decision rules are adapted from decision theory, e.g. expected utility maximization. Paired with accuracy as the relevant measure of utility, we end up with the decision rule:

Maximize Expected Accuracy: adopt the credence function that has the highest expected accuracy, by your own lights.[15]

The expected accuracy of a credence function is standardly calculated as the sum of a credence function's accuracy in each world, weighted by the probability of that world. In symbols:

$$EU^{Cr}(Cr') = \sum_{w_i \in \mathscr{W}} Cr(w_i) U(Cr'', w_i)$$

With the decision rule Maximize Expected Accuracy, various results can be proven. Call a function from evidence to credence functions an 'update policy.' Greaves & Wallace (2006) and Leitgeb & Pettigrew (2010) proved that from an agent's own perspective, given the choice of all possible update policies, Conditionalization alone maximizes expected accuracy. So, one might conclude hastily, in order to be an ideal information gatherer, your credences should update by Conditionalization.

But, you might ask, isn't example #4 intuitively a case where I know that some other credences than my own maximize expected accuracy from my point of view?

---

[15] (Carr manuscript) argues against the conception of expected accuracy used by epistemic utility theorists. For the purposes of addressing this objection to diachronic rationality, though, I will take the appeal to expected accuracy at face value.

In that example, I would receive an increase in expected accuracy if I updated by some means that would involve violating Conditionalization. Does that example conflict with the results of epistemic utility theory?

4.3 Assessing rational trade-offs

In fact, there's no conflict between the idea that there could be rational information trade-offs (violating Conditionalization) and the epistemic utility theoretic result that Conditionalization is the only update policy that maximizes expected utility.

The reason: update policies are narrowly defined as functions from evidence to particular credence functions. But it's a feature of information trade-offs that you do not know, in advance, what credences you will adopt as a result of taking the trade-off. (If you did, then you could update on that information directly, which would then amount to pure information gain.) Indeed, on common assumptions,[16] it cannot be the case for any particular credence function that you can rationally assign it higher expected accuracy than your own credence function. But if you have the option of adopting *whichever* of a set of possible credence functions that has updated on some information (information that is not otherwise accessible to you), then that option can maximize expected accuracy from your perspective.

Let's consider a particular case of an informational trade-off, specifying some of the details from example #4. Suppose a particular coin is either fair or biased (with a $\frac{3}{4}$ heads bias), and it will land either heads or tails. You are uncertain about both matters. Now, you and your colleague start with the same priors:

$$w_{FH}: \text{fair, heads} \quad Cr_0(w_{FH}) = \tfrac{1}{4}$$

$$w_{FT}: \text{fair, tails} \quad Cr_0(w_{FT}) = \tfrac{1}{4}$$

$$w_{BH}: \text{biased, heads} \quad Cr_0(w_{BH}) = \tfrac{3}{8}$$

$$w_{BT}: \text{biased, tails} \quad Cr_0(w_{BT}) = \tfrac{1}{8}$$

Then you learn whether the coin lands heads or tails. Your colleague learns whether the coin is fair or biased. Both of you conditionalize on your respective evidence. You are not permitted to know the answers to both questions.

Suppose you learn that the coin lands heads. You have a credence downloader that will allow you to perform the informational trade-off. Is it epistemically rational for you to give up your knowledge in order to gain your colleague's?

Applying the rule Maximize Expected Accuracy isn't straightforward. Since we don't know what your colleague has learned, we don't know which credence

---

[16] Namely, that epistemic utility functions must be *proper* in the sense that they yield the result that any coherent credence function maximizes expected accuracy by its own lights.

function to assess. So it's not obvious how we can even determine the expected accuracy of your colleague's credence function.

Here is my suggestion: we can introduce a new kind of epistemic action. Call it *learning the answer to a question.* Learning the answer to a question involves taking an epistemic option when you're not in a position to know what credence function it will result in your adopting.[17]

For a question $\mathscr{Q}$ (i.e. a partition over the set of epistemically possible worlds), let $Cr_{\mathscr{Q}}$ be Cr_0 conditionalized on whatever the true answer to $\mathscr{Q}$ is (that is, whichever proposition in $\mathscr{Q}$ is true at the world of assessment).

In our example, we can call whatever credence function your colleague has after learning whether the coin is biased or fair $Cr_{\mathscr{Q}_{BF}}$. Note that '$Cr_{\mathscr{Q}_{BF}}$' is a description: it picks out different credence functions in different worlds. Ex hypothesi, your colleague updates on B in B-worlds and on F in F-worlds.

Now, with a concrete example in hand, and a new tool (the epistemic act of learning the answer to a question), we can ask: should you take the trade-off? We need to explain how to calculate the expected accuracy of $Cr_{\mathscr{Q}_{BF}}$ from your point of view:

1. Calculate the accuracy of Cr_B at B-worlds and Cr_F at F-worlds.

2. Sum the values, weighted by their probability according to Cr_H.

In symbols:
$$EU^{Cr_H}(Cr_{\mathscr{Q}_{BF}}) = \sum_{w_i \in \mathscr{W}} Cr_H(w_i)U(Cr_{\mathscr{Q}_{BF}}, w_i)$$

In this case, with plausible assumptions about the accuracy function U, taking the trade-off maximizes expected accuracy. Retaining your current credences does not.[18]

---

[17] This kind of epistemic tool isn't just for science fictional cases where you are offered information trade-offs. We can do other things with our new epistemic acts. For example, they can be useful in decisions over whether it would be more informative to perform one experiment or another, in circumstances where it is impossible, or at least costly, to perform both.

[18] Suppose U is the negative Brier Score. (Joyce (2009) and Leitgeb & Pettigrew (2010) argue that only scoring rules like the Brier Score satisfy plausible constraints on epistemic utility functions.) Where $v_w(X) = 1$ if X is true at w and $v_w(X) = 0$ if X is false at w, $U(p, w) = -\sum_{X \subseteq \mathscr{W}} |v_w(X) - p(X)|^2$.

$$\begin{aligned}
EU^{Cr_H}(Cr_{\mathscr{Q}_{BF}}) &= \sum_{w_i \in \mathscr{W}} Cr_H(w_i)U(Cr_{\mathscr{Q}_{BF}}, w_i) \\
&= Cr_H(w_{BH})U(Cr_B, w_{BH}) + Cr_H(w_{HF})U(Cr_F, w_{FH}) \\
&= -\tfrac{11}{40}
\end{aligned}$$

This isn't surprising. Knowing that the coin landed heads isn't particularly informative about whether the coin is fair or biased, since it would be unsurprising either way. On the other hand, if you had instead learned that the coin had landed tails, then it would maximize expected accuracy to reject the trade-off. After all, knowing that the coin landed tails gives you fairly strong evidence in support of the coin's being fair.

So, we have a concrete case where taking an informational trade-off maximizes expected accuracy.

4.4 Discussion

As I said before, the defender of diachronic norms had two options for responding to an objection like this. If she continues to endorse Conditionalization, then she must reject the claim that it's rational to accept informational trade-offs. (This might involve rejecting the idea that we should perform those epistemic acts that maximize expected accuracy, or it might involve rejecting the idea that taking an information trade-off is an appropriately understood as an epistemic act.)

On the other hand, if we allow informational trade-offs as epistemic options, then accepting trade-offs can lead to maximizing expected accuracy. And if we accept that this is rational, then we can replace Conditionalization with amore liberal diachronic rule.

These two options provide us with different pictures of what an ideally rational agent's credences will look like over time. On the Conditionalization picture, the ideal rational agent's stock of information will only ever increase. But if we allow for violations of Conditionalization in informational trade-offs, then the ideally rational agent will in some circumstances take epistemic risks. These risks have two salient features that distinguish them from obeying Conditionalization. First, they involve sure loss of information; second, they may lead to decreases the agent's expected accuracy (from the perspective of her updated credences).

$$
\begin{aligned}
EU^{Cr_H}(Cr_H) &= \sum_{w_i \in \mathscr{W}} Cr_H(w_i)U(Cr_H, w_i) \\
&= Cr_H(w_{BH})U(Cr_H, w_{BH}) + Cr_H(w_{HF})U(Cr_H, w_{FH}) \\
&= -\tfrac{12}{25}
\end{aligned}
$$

The expected accuracy of your colleague's credence function ($Cr_{\mathscr{Q}_{BF}}$) is greater than the expected accuracy of your own credence (Cr_H). So if you know H, expected accuracy maximization requires you to take the trade-off.

Here is a candidate liberal diachronic norm (which is a variant on diachronic evidentialism):

Liberal norm: An ideally rational agent's credences change only in order to maximize their expected accuracy.

Note that for cases of pure information gain, Conditionalization will still hold. Furthermore, rational trade-offs arguably only occur in sci-fi cases.[19] So, in ordinary cases, a more traditional, strict norm will still hold.

Strict norm: An ideally rational agent's credences only change in response to new evidence.

5   Conclusion

I've argued that there is a conflict between diachronic norms of epistemic rationality and a form of epistemic internalism. I've also argued that diachronically coherent agents are epistemically better. We should think of epistemic rationality as providing constraints that allow us to be more informed about our environment, whatever our environment happens to be like.

The diachronic norms I've advocated are at a middle ground between epistemic internalism and externalism: they are sensitive to facts that are external to the time-slice, but not necessarily external to the person. Contrast this sort of view with process reliabilism, which is concerned with whether some belief-forming process *actually* conduces toward the truth. Whether it does will depend on contingent facts about the agent's environment. A norm like expected accuracy maximization is concerned with whether an update method is *likely* to conduce toward the truth, by the believer's own lights.

If we take the option of maintaining Conditionalization, we are also given at a middle ground between epistemic conservatism and evidentialism. Like conservatism, Conditionalization permits us to continuing to believe a proposition if we already believe it (with certainty). In fact, Conditionalization requires it. But unlike conservatism, Conditionalization doesn't permit continuing to believe a proposition after the evidence for it has been forgotten. Conditionalization requires remembering the evidence as well. In short, Conditionalization doesn't permit violations diachronic evidentialism. Hence, what we're required to believe is always determined by what our evidence supports. It's just that our evidence—what we've learned—might escape us.

---

[19] One might make the case that clutter avoidance is a more psychologically realistic version of an informational trade-off; see (Harman 1986).

References

Arntzenius, Frank (2003). "Some Problems for Conditionalization and Reflection." *The Journal of Philosophy*, 100(7): pp. 356–370.

Burge, Tyler (1997). "Interlocution, perception, and memory." *Philosophical Studies*, 86: pp. 21–47.

Carr, Jennifer (manuscript). "How to Expect When You're Expecting."

Christensen, David (2000). "Diachronic Coherence Versus Epistemic Impartiality." *Philosophical Review*, 109(3): pp. 349–371.

Friedman, Jane (manuscript). "Inquiry and Agnosticism."

Greaves, Hilary, and David Wallace (2006). "Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility." *Mind*, 115(459): pp. 607–632.

Harman, Gilbert (1986). *Change in View: Principles of Reasoning*. MIT Press.

Hedden, Brian (2013). "Options and Diachronic Tragedy." *Philosophy and Phenomenological Research*, 87(1).

Joyce, James (1998). "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science*, 65(4): pp. 575–603.

Joyce, James (2009). "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief." In *Degrees of Belief*, Synthese, vol. 342, pp. 263–297.

Leitgeb, Hannes, and Richard Pettigrew (2010). "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy." *Philosophy of Science*, 77(2): pp. 236–272.

McGrath, Matthew (2007). "Memory and Epistemic Conservatism." *Synthese*, 157(1): pp. 1–24.

Meacham, Christopher J. G. (2010). "Unravelling the Tangled Web: Continuity, Internalism, Non-Uniqueness and Self-Locating Beliefs." In *Oxford Studies in Epistemology, Volume 3*, Oxford University Press.

Moss, Sarah (manuscript). "Credal Dilemmas."

Talbott, W. J. (1991). "Two Principles of Bayesian Epistemology." *Philosophical Studies*, 62(2): pp. 135–150.

Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.