Bayesian Coherentism

(penultimate draft—please cite published version)

This paper considers a problem for Bayesian epistemology and proposes a solution to it. On the traditional Bayesian framework, an agent updates her beliefs by Bayesian conditioning, a rule that tells her how to revise her beliefs whenever she gets evidence that she holds with certainty. In order to extend the framework to a wider range of cases, Richard Jeffrey (1965) proposed a more liberal version of this rule that has Bayesian conditioning as a special case. Jeffrey conditioning is a rule that tells the agent how to revise her beliefs whenever she gets evidence that she holds with any degree of confidence. The problem? While Bayesian conditioning has a foundationalist structure, this foundationalism disappears once we move to Jeffrey conditioning. If Bayesian conditioning is a special case of Jeffrey conditioning, then they should have the same normative structure. The solution? To reinterpret Bayesian updating as a form of diachronic coherentism.

1 Introduction

Foundationalism and coherentism are competing views about the structure of epistemic justification. It's surprising, then, that they co-exist on the Bayesian framework. The explanation: Bayesianism is committed to norms that govern our degrees of belief—our credences—in propositions that stand in particular logical relations to each other at each time. It's also committed to norms that govern how these credences change over time in response to new evidence. Traditional Bayesian epistemology is coherentist with respect to the first set of norms. It's foundationalist with respect to the second. It has two strands of justification running through it.

This paper considers a problem for Bayesianism's second strand of justification and proposes a solution to it. On the traditional Bayesian framework, an agent updates her beliefs by Bayesian conditioning, a rule that tells her how to revise her beliefs whenever she gets evidence that she holds with certainty. In order to extend the framework to a wider range of cases, Richard Jeffrey (1965) proposed a more liberal version of this rule. Jeffrey conditioning is a rule that tells the agent how to revise her beliefs whenever she gets evidence that she holds with *any* degree of confidence. Jeffrey claimed that his rule has Bayesian conditioning as a special case. This claim is now a truism of Bayesian epistemology.

The problem? While Bayesian conditioning has a foundationalist structure, this foundationalism disappears once we move to Jeffrey conditioning. But if Bayesian conditioning is a special case of Jeffrey conditioning, then these two updating rules should have the same normative structure. We are, then, left with the following inconsistent triad: (1) If one norm is a special case of another, then they should have the same normative structure, (2) Bayesian conditioning is a special case of Jeffrey conditioning, (3) Bayesian conditioning and Jeffrey conditioning have different normative structures.

In this paper, I'll argue for an interpretation of the Bayesian framework that resolves this inconsistency by rejecting (3). I'll argue that both regular Bayesian updates and Jeffrey updates proceed from a common framework—one with a coherentist structure. My strategy will be to appeal to what has long been deemed to be a defect of Jeffrey conditioning: the fact that its updates aren't guaranteed to commute. To say that Jeffrey updates aren't guaranteed to commute is to say that an agent's credences after a sequence of updates will sometimes be determined by the order in which this evidence has been received. This feature of Jeffrey updates is a defect because the order in which some evidence has been received seems irrelevant to the impact it ought to have. While the fact that the Jeffrey framework can't guarantee that its updates will commute is standardly taken to show that the framework fails to satisfy an important desideratum for an updating rule, in this paper, I propose that we take the commutative property to play a more fundamental role. I propose that we take the commutative norm that Bayesianism is committed to to be one that *grounds* particular updates. In other words, some set of updates will be justified *to the extent that they commute*. Since the sort of consistency this norm encodes is to updates what the norm of evidential consistency from traditional formulations of coherentism is to beliefs, it looks as though the best way of understanding the structure of Bayesian updating is not as a form of diachronic foundationalism, but as a form of diachronic coherentism.

Before we get started, let me say a bit more about our inconsistent triad. The claim that Bayesian conditioning and Jeffrey conditioning have different normative structures will be argued for below. And the claim that Bayesian conditioning is a special case of Jeffrey conditioning is a mathematical fact, as we will also see in just a moment. Before we move on, however, I want to briefly defend (1): the claim that if one norm is a special case of another, then they should have the same structure. They should function in the same way. It might seem as though counterexamples to this idea aren't difficult to find. Consider a rule that tells you that you are permitted to drive in the carpool lane if you have any number of people in the car—except for just one. Here the special case of having just one person in the car calls for something different from the general case of having any number of people in the car. But this doesn't seem unreasonable. Or consider a law that requires a unanimous guilty verdict from a jury in order for the death penalty to be imposed. Though this law might be controversial, again, it's not clear this is due to the fact that it treats the special case of unanimity as different from the more general case of a jury reaching its decision.[1]

It's true that, when taken on its own, the special case does not seem unreasonable in either of these examples. But I think there's at least a sense in which, considered side by side with an instance of the more general case, the special case calls for an explanation. In both of these examples, it seems reasonable to question the sharp cut-off between the special case and the one just before it. It seems reasonable to ask why a car with only one person in it is so different from a car with two people in it, and why a unanimous jury is different in kind from a jury with only one dissenter.

In the same way, I want to suggest that the fact that the special case of Bayesian condi-

[1] Thanks to Jonathan Weisberg for this last example.

tioning has a certain normative structure, while not problematic when considered on its own, is odd when considered together with the fact that Jeffrey conditioning, in general, does not have this same structure. The oddness in all of these cases may be defeasible. If one were to explain that unanimous agreement is necessary to establish that there is no reasonable doubt about a defendant's guilt, one might then see what makes a unanimous jury qualitatively different from a jury with one dissenter. In the same way, there are assumptions we might make about the Bayesian formalism that would justify treating updates on certain evidence as qualitatively different from updates in general. In the last section of this paper, I'll briefly consider some of these assumptions. But, for now, I'll assume there's at least something to be said for an interpretation of Bayesian updating that yields the result that certain and uncertain updates proceed from a framework with the same normative structure.

One more thing before moving forward. Since the Jeffrey framework includes both updates on certain and uncertain evidence, one might object that there is no difference in normative structure between the Jeffrey framework and the regular Bayesian framework. Jeffrey conditioning is constrained in the very same way as Bayesian conditioning: like Bayesian conditioning, Jeffrey conditioning imposes a foundationalist constraint on evidence that receives a credence of one. However, the interesting question is not whether the Bayesian framework and the Jeffrey framework include the same constraint. The interesting question is whether this constraint governs *all* of the updates that proceed from each of these frameworks. The interesting question is whether the constraint that governs updates made on evidence that we hold with certainty also governs updates made on evidence that we hold with any degree of confidence. On the orthodox way of understanding Bayesian updating, it does not. And this seems objectionable. If the change in the values that our evidence receives is meant to represent a change in the certainty of our evidence, this change in value should not entail a difference in anything else.

Here's how the discussion will go. In §2, I classify different approaches to Bayesian updating in a way that makes space for the constraint I go on to defend. In §3, I describe the sense in which regular Bayesian conditioning has a foundationalist structure. In §4, I explain why adopting Jeffrey conditioning entails abandoning this foundationalism. In §5, I propose a constraint that makes Bayesian updating look like a form of coherentism and argue that this constraint can make better sense of the truism that Bayesian conditioning is a special case of Jeffrey conditioning. In §6, I give, in broad strokes, the formal details of this constraint. In §7, I consider this constraint in action. Finally, in §8, I revisit the motivation for this discussion.

2 Diachronic Coherence for Bayesians

2.1 *The Normative Approach*

I've suggested that it's possible to ask whether Bayesian updating is a form of foundationalism or a form of coherentism. Let's begin by getting clear on exactly what this question means.

Standard Bayesianism assumes that an agent's credences in the propositions she entertains can be represented as an assignment of real numbers to those propositions. It further assumes that two norms of coherence govern this assignment. First, Bayesianism is committed to the constraint that, at each time, an agent's credences be a probability function. To say that a Bayesian agent is synchronically coherent is to say that she conforms to Probabilism. Second, Bayesianism is committed to the constraint that an agent's beliefs evolve over time in accordance with her conditional probabilities. To say that a Bayesian agent is diachronically coherent is to say that the values of the agent's probabilities, conditional on her evidence, remain the same before and after she revises her beliefs. We can think of the agent's conditional probabilities as arrows that proceed from her evidence and that guide the propagation of the rest of her probabilities. To serve this guiding function, they must remain fixed.[2]

As it happens, every probabilistic belief transition has a set of arrows.[3] For any probabilistic belief transition, there will be *some* information that each of my beliefs are conditional on in the same way before and after this transition. More formally, for every probabilistic belief transition, there will always be a partition of propositions (a set of mutually exclusive and exhaustive propositions) that the agent's doxastic state is defined over, and that is sufficiently fine-grained to enable us to describe the transition in question as conditional on this partition:

> **The Descriptive Account:** There is a sufficient partition, $\{B_i\}$, for every probabilistic belief transition.[4]

Or, equivalently, where $S=\{B_1, ..., B_n\}$ is a partition of propositions, and where an agent has an experience that causes her to revise her beliefs, the transition between the agent's prior probability distribution, p, and posterior probability distribution, p', can be formulated in a way that underlines that there will always be *some* partition that can be described as guiding her belief revision:

> **The Descriptive Account:**
> $\forall p \forall p' \exists S (\forall B_i \in S) \forall A (p(A|B_i) = p'(A|B_i))$, if defined.

[2] The arrow analogy is borrowed from Weisberg (2015). This guiding feature of our conditional probabilities is often referred to as 'rigidity' (see Jeffrey (1965)).

[3] A couple of notes about terminology. First, I will for the most part use the expressions 'belief transitions' and 'updates' interchangeably. Notably both expressions will sometimes be used in contexts where their standing as diachronically coherent is what is at issue. Relatedly, I will use the term 'evidence' a little ambiguously throughout. For instance, I will talk about 'uncertain evidence' in contexts where whether uncertain information meets the relevant normative constraint on evidence is, again, the question we are trying to decide. While the ambiguity here isn't ideal, context should make clear enough what is being intended.

[4] For the proof of this, see Blackwell and Girshick (1979, p. 218) and Diaconis and Zabell (1982, p. 824). As Diaconis and Zabell note, there will be cases where our conditional probabilities are undefined for some partition—namely, where we assign a member of our partition a credence of zero. However, their result still holds for all updates if we take a sufficient partition to be a partition that is sufficient to represent a probabilistic belief transition as an update that is conditional on every proposition in this partition for which a conditional probability is defined (*Cf.* Blackwell and Girshick (1979, §8.4.3)).

Since the Descriptive Account holds for any two probability distributions, it is no stronger than Probabilism. Since it is no stronger than Probabilism, one might worry that it will be too weak to capture any interesting notion of diachronic coherence.

We can remedy this by strengthening our account of diachronic coherence. We can do this by stipulating that it is only when an agent conditions her beliefs on a partition that satisfies some additional normative constraint on evidence that she is diachronically coherent:

> **The Normative Approach:** A probabilistic belief transition is always such that,
>
> (a) there is a sufficient partition, $\{B_i\}$, for this transition, and
>
> (b) an agent who undergoes this transition is diachronically coherent iff $\{B_i\}$ satisfies some normative constraint.

Or, equivalently, where S=$\{B_1, ..., B_n\}$ is a partition of propositions whose values satisfy some normative constraint, and where, p, is the agent's prior probability distribution, and, p', is the agent's posterior probability distribution, an agent is diachronically coherent just in case the following holds:

> **The Normative Approach:**
> $\forall p \forall p' (\forall B_i \in S) \forall A (p(A|B_i) = p'(A|B_i))$, if defined.

Since not every probabilistic agent will revise their beliefs in accordance with whatever constraint on evidence we use to fill in (b), the Normative Approach avoids the worry that it is too weak to capture any interesting notion of diachronic coherence.

While the Normative Approach seems plausible, it has, to the best of my knowledge, never been used to develop an account of diachronic coherence. However there is a way of understanding diachronic coherence for Bayesians that is more orthodox and that is also stronger than the Descriptive Account. This understanding carves a middle path between descriptive and normative approaches to diachronic coherence, and is perhaps the understanding of diachronic coherence that is most widely assumed. This way of thinking about diachronic coherence says that a Bayesian agent's diachronic obligations consist in how she ought to proceed *supposing* she has gotten some weighted partition as evidence:

> **The Orthodox Account:**
> Where S= $\{B_1, ..., B_n\}$ is a partition of propositions whose values have been changed directly by experience, and where, p, is the agent's prior probability distribution, and, p', is the agent's posterior probability distribution, an agent is diachronically coherent just in case the following holds:
>
> $\forall p \forall p' \forall B_i (\exists S (B_i \in S) \rightarrow \forall A (p(A|B_i) = p'(A|B_i)))$, if defined.[5]

[5] For some canonical examples of the Orthodox Account, see Skyrms (1987), Howson and Urbach (1989) and Jeffrey (1992), among many others.

Since not every probabilistic agent will, in virtue of being a probabilistic agent, revise their beliefs in accordance with the particular evidence they have, this approach avoids the worry that it is too weak to capture any interesting notion of diachronic coherence. But since this understanding of diachronic coherence does not require the agent's evidence to satisfy any additional normative constraint, it is also not as strong as the Normative Approach.

I want to defer saying anything more about the Orthodox Account for the moment. It will become clear a bit later on why this account of diachronic coherence cannot be used to unify Bayesian updates in the way that we are looking to do.

2.2 *A Formal Account of Evidence*

For now, then, let us assume that the aim of this paper will require that we adopt the Normative Approach to diachronic coherence. And this will require that we adopt an account of evidence. There are a couple of ways that we might go about this. The most familiar of these ways is to appeal to a substantive account of evidence; for instance, to the requirement that evidence be what one knows, or be related to what one has internal access to, or be formed by a reliable process, etc. What makes these accounts substantive ones is that Bayesians, qua Bayesians, aren't committed to the normativity of knowledge, or of access, or of reliability, etc. A substantive account of evidence is a constraint on evidence formulated in terms of a property that is not entailed by the Bayesian formalism.

This paper will take a different approach by defending a formal account of evidence. An account of evidence is formal just in case it's not substantive; just in case it *is* formulated uniquely in terms of some feature of the Bayesian formalism. A formal account of evidence will hold that it is in virtue of the sufficient partition of an update being assigned certain values, or weights, that the agent can be said to have evidence, where these values are not further justified by the sorts of substantive considerations just mentioned.[6] Exactly what it will look like for a formal constraint on evidence to be satisfied will become clearer in the next section. For now, however, notice that the appeal to a formal account of evidence leaves us able to understand how it is possible to ask whether Bayesian updating is a form of foundationalism or a form of coherentism. Since what we will be after is a formal, or structural, account of evidence, and since foundationalism and coherentism are both structural norms, they will both be candidates for such an account.

Why defend a formal account of evidence in the first place? Because I think it's of interest to consider how much normativity can be defined of in terms of the commitments Bayesians already hold. It's worth noting that my formal account of evidence will entail

[6] Perhaps an easier way of understanding what makes knowledge and access and reliability substantive, rather than formal, requirements is that they can be made sense of out of context. They can be defined independently of the other epistemic commitments one happens to hold. Formal norms are different in this regard. Take, for instance, the formal norm of consistency. The way that a Bayesian defines consistency will differ from the way that a defender of Dempster-Shafer theory does this. While the Bayesian will spell out her notion of consistency by means of probability functions (e.g., "X is consistent if it treats these probability functions in this way"), a Dempster-Shafer theorist, who trades in belief functions, will define her notion of consistency in terms of those. Unlike knowledge or access or reliability, the norm of consistency is so thin that it isn't complete without a framework already in place to fill it in.

an account of diachronic coherence with a distinguishing feature. On the account I end up defending, diachronic coherence isn't defined for an individual belief revision. Instead, it is a relation that is defined over *sets* of belief revisions.[7] We've said that an agent's standing as synchronically coherent will depend upon how her credences are related to each other. My account entails that an agent's standing as diachronically coherent will depend upon how her belief revisions are related to each other. My account leaves us with an interpretation of Bayesian epistemology that is coherentist with respect to both strands of justification that run through it.

2.3 *An Assumption*

Finally, an assumption. Epistemic theories can be given one of two interpretations. On the one hand, we might think that what any such theory provides is guidance for how a rational agent ought to act. On the other hand, we might think that what any such theory provides is a way of evaluating an agent's actions, or the state of affairs that this action brings about, whether or not we would want to say that an agent *ought* to have done what she did. Bayesian epistemology, when understood in the first of these ways, has received its fair share of criticism. This is because the sorts of idealizing assumptions that we need to get the framework off the ground require of ordinary agents that they perform operations that are computationally intractable. Here's Harman (1986, p. 25-26) on this:

> One can use conditionalization to get a new probability for *P* only if one has already assigned a prior probability not only to *E* but to *P*⊠*E*. If one is to be prepared for various possible conditionalizations, then for every proposition *P* one wants to update, one must already have assigned probabilities to various conjunctions of *P* together with one or more of the possible evidence propositions and/or their denials. Unhappily, this leads to combinatorial explosion, since the number of such conjunctions is an exponential function of the number of possibly relevant evidence propositions.

And Earman (1992, p. 56):

> 'Ought' is commonly taken to imply 'can', but actual inductive agents can't, since they lack the logical and computational powers required to meet the Bayesian norms. The response that Bayesian norms should be regarded as goals toward which we should strive even if we always fall short is idle puffery unless it is specified how we can take steps to bring us closer to the goals.

In light of these sorts of criticisms, I will assume that Bayesianism is best understood as a set of evaluative norms rather than as a set of action-guiding norms. This means that although Bayesian epistemology sets certain standards, there are no obligations issued

[7] A consequence of this is that the Normative Approach is more perspicuously formulated in terms of a set of evidence partitions, instead of just one. How this can be done will become clearer in §5.

by the theory. Just as we can say that cars are good insofar as the brakes work, and bad insofar as they don't, without imposing any obligations on anyone to do anything, we can say that updates are good or bad, in virtue of certain features of them, without imposing any obligations on anyone to do anything.

Notice that the way we've set things up in this section already encourages us to think of Bayesian epistemology as an evaluative theory. The natural question to ask on the action-guiding approach is: given what I take my evidence to be, how should I update? By contrast, the natural question to ask on the evaluative approach is: does my update have the right features? By defining evidence *in terms* of the update that it triggers, rather than vice versa, we set ourselves up to pursue the second of these questions. An agent's update will be good insofar as the formal constraint on evidence defended in this paper is satisfied, and bad insofar as it isn't. However this does not guide the agent, or obligate her to update, in any particular way.[8]

3 Bayesian Foundationalism

I've claimed that Bayesian conditioning can be understood as a form of foundationalism about updating.[9] To better understand the sense in which Bayesian conditioning has sometimes been claimed to be a form of foundationalism, it will help to first get clear about what foundationalism amounts to in its more traditional guise—as a structure that applies to beliefs.

Traditional foundationalism about epistemic justification says that the ultimate source of the justification of all our beliefs is some privileged set of cognitive states that is the locus of this justification, but that can't be the target of it. It's the conjunction of the claims that (1) some cognitive states are basic, in the sense of their being justified not in virtue of their relations to other cognitive states, and that (2) all non-basic states are justified in virtue of some relation that they bear to basic states. In addition, classical foundationalism assumes that (3) the distinguishing mark of basic states is their infallibility.[10]

As this description makes clear, foundationalism is *just* a structure. It takes no stand on what features some cognitive state must have in order to be a basic state, other than the role that it occupies on the foundationalist's picture of things. Why think that Bayesian conditioning is a form of foundationalism? Consider the way our rough definition of

[8]Despite this, I will continue to speak more loosely in terms of 'norms' and 'oughts'. I do this mainly for ease of exposition, but also because it seems plausible that this language can be appropriately applied, in an evaluative way, to states of affairs. (For this point, see, for instance, Chrisman (2008)).

[9]I continue to take 'Bayesian conditioning' to refer, very generally as in §1, to Bayesian updating on certain evidence. Those who have explicitly taken Bayesian conditioning to instantiate a foundationalist structure include Christensen (1992), Bradley (2005), Weisberg (2009) and Titelbaum (forthcoming), among others, though these discussions do not necessarily take this foundationalism to provide the sort of formal account of evidence I have in mind.

[10]By contrast, many non-classical foundationalist accounts, like Goldman (1988)'s reliabilism, Plantinga (1991)'s proper basicality, Pryor (2000)'s dogmatism and Huemer (2007)'s phenomenal conservativism defend some form of fallible foundationalism. Each of these accounts maintains that the property that makes beliefs basic is something other than their infallibility. We'll return to consider the relevance of these accounts in the next section.

Bayesian conditioning slots into the schema from above:

> **Bayesian Conditioning:** A probabilistic belief transition is always such that,
>
> (a) there is a sufficient partition, $\{E_i\}$, for this transition, and
>
> (b) an agent who undergoes this transition is diachronically coherent iff some E_i is held with certainty.

Where we assume the sort of foundationalism that identifies a certain belief with an infallible one, Bayesian conditioning satisfies (3) by requiring that the agent's evidence be a proposition of which she is certain.[11] The previous formulation clearly satisfies (2) as well: the values we assign the rest of our beliefs depend upon our evidence qua basic state. What about (1)? While the agent uses her evidence proposition to infer the credences she holds in other propositions, the evidence proposition itself cannot receive this same sort of support. This is because propositions that receive a credence of one cannot have their values changed by Bayesian conditioning at some later time. That certainties stay certainties is a mathematical feature of the Bayesian formalism. Crucially, then, *once* a belief becomes a basic state—once it becomes evidence—it is no longer able to receive the same sort of inferential support that it offers.[12]

In short, then, traditional foundationalism makes justification a function of whether some belief (P) is in the set of beliefs justified by the agent's basic state (S):

> **Traditional Foundationalism**: f_S: P $\mapsto \{0, 1\}$,
>
> (1 if p $\in$ S, and 0 otherwise)

In a similar way, we can understand Bayesian conditioning as the constraint that some update be in the set of updates that is justified by the agent's basic state, i.e., her evidence (E):

> **Bayesian Foundationalism**: f_E: UP $\mapsto \{0, 1\}$,
>
> (1 if up $\in$ E, and 0 otherwise)

[11] One might object that to identify a certain belief with an infallible one is to make precisely the sort of substantive assumption that a formal account of evidence is supposed to avoid. However, this assumption seems less divisive than the ones that we appealed to earlier as paradigmatically substantive assumptions, e.g., that internalism or externalism is the correct theory of justification, etc. Moreover, if one is uncomfortable with this assumption, one can do without it since, as we've already said, (3) is not a necessary component of a foundationalist account. As I will go on to argue in the next section, it doesn't matter much how we think about the property that makes a state foundational. In assuming that the property that makes some state a foundation is certainty, I am simply following other Bayesians who have made this assumption (see fn. 9).

[12] I take this to be consistent with the idea that *before* a belief becomes evidence—when it is merely a proposition that the agent has some credence in—it is able to receive some inferential support.

Like traditional foundationalism, Bayesian foundationalism begins from the assumption that although foundationalism does not entail a substantive account of evidence, it can nevertheless tell us about the structure that would guide this evidence, if it were there. And we can still say that this structure has a good-making feature. A Bayesian update on certain evidence has the formal, good-making feature of forever eliminating any beliefs that are inconsistent with those beliefs that we hold with certainty. This structure is Bayesianism's formal constraint on evidence, and it defines the sense in which Bayesian conditioning is often assumed to be a form of foundationalism.

4 Foundationalism Undermined

Most take the fundamental idea behind Jeffrey conditioning to be the thought that, as Jeffrey (1983, p.171) put it, "it is rarely or never that there is a proposition for which the direct effect of an observation is to change the observer's degree of belief in that proposition to one." Most of the time we have an experience that changes our credence in some proposition without making us sure of it. We get a quick glimpse of color on the floor that makes us think that the sock might be red. But maybe it's really brown. Or maybe it's purple.

To capture this more realistic class of cases, we need a rule that tells us how we should revise our beliefs whenever we get this sort of uncertain evidence. Jeffrey (1965) introduces a rule that does just this by assuming that our evidence is a partition of propositions that can be assigned values other than zero and one. Formally, Jeffrey conditioning has Bayesian conditioning as a special case. Both updating rules say that we should revise our beliefs in accordance with the conditional probability that our evidence determines. Assuming our evidence to be a partition allows us to accommodate the uncertainty of some pieces of evidence by allowing us to assign probabilities other than one and zero to the possibility that the sock is red, and to the possibility that it is brown, and to the possibility that it is purple—which, together, will sum to one. Assuming our evidence to be a partition also allows us to accommodate the certainty of some pieces of evidence by allowing us to assign probability one to the possibility that the sock is red and probability zero to the possibility that it isn't.

But although Bayesian conditioning is a special case of Jeffrey conditioning, it is a degenerate special case of it. The notion of degeneracy that I have in mind is the one that is familiar from mathematics. Sometimes what we get from the limiting case of a class of objects is a change in one of its usual parameters. An example is the way that a line segment is a degenerate case of a triangle: it is the triangle one obtains when one angle has a value of 180 degrees and the other angles have values of zero. Something similar holds where we think of Bayesian conditioning as the special, limiting case of Jeffrey conditioning. Formally it meets all the criterion for being such a case. But it is a case that adds normative structure to the Jeffrey framework in the way that a line segment abstracts away one of the dimensions of a triangle.

Does this analogy get the Bayesian off the hook? Earlier I suggested that the claim that all cases of Bayesian updating should have the same structure may be defeasible. Having suggested that it is pretty common in mathematics for the special, limiting case of a class of objects to fail to have the same structure as the more general case, one might argue that we now have quite strong reason to think that this earlier reasoning *has* been defeated. Maybe we even have reason to think that this reasoning was mistaken all along.[13]

I think this is the wrong conclusion. While the mathematics example provides a useful analogy for what is going on in the case under discussion, it's not an analogy that justifies this case. To justify treating the special case of a norm differently from the more general case, we need to establish not merely that there exists a plausible way of representing the different natures of these cases, but that there exists a plausible way of *justifying* this difference. We need a normative reason for this difference. Recall the example of the jury. We said earlier that what would justify treating a unanimous jury differently from a jury with only one dissenter is the assumption that only the former kind of jury could eliminate all reasonable doubt. The fact that only a unanimous jury could eliminate all reasonable doubt is a normative reason for treating this special case differently. And it is clearly the sort of reason that we should be interested in. Or consider again the case of the carpool lane. In order to justify the more-than-one-person-in-the-carpool-lane rule, it won't do simply to point out that there is a mathematical structure that is able to capture a difference between the case where there is one person in the car and the case where there are some number of people in the car. Instead we need to say something about why this structure should be adopted. We need to say something about why there *should* be a sharp cut-off point between the special case and the one just before it that justifies treating these cases in different ways. Maybe we can. Maybe the reason the carpool lane should be able to be occupied by a car with two people in it, but not one, is because this is the minimal restriction required to reduce pollution to a level where it is not harmful. This is the kind of justification that illustrates the defeasibility of the claim that the special case (of having one person in the car) and the general case (of having some number of people in the car) ought to be treated the same.

In short, a normative difference requires a normative explanation. Near the end of the paper, I'll briefly consider the sort of normative explanation that might justify the degeneracy of the special case of Bayesian conditioning. For now, I'll simply note that the mere prevalence of a class of mathematical objects isn't capable of doing this.

It's a problem, then, that Bayesian conditioning includes a foundationalist constraint that updates on uncertain evidence lack. Moreover, there is more than one way of understanding how updates on uncertain evidence lack this constraint. First, assume that we take the agent's basic state to be the partition that she updates on. Since the propositions in this partition can receive a value of less than one—less than complete certainty—this partition does not include an infallible belief. More importantly, since the propositions in this partition can receive a value of less than one, they are able to have their values changed by means of the same sort of inferential support that they offer by a future update. Therefore, these partitions violate the first and third conditions of foundationalism identified above.

Many non-classical foundationalists also reject the first and third conditions identified above. Goldman (1988)'s reliabilism is an example of an externalist account that

[13] Thanks to an anonymous referee for raising this objection.

does this. Pryor (2000)'s dogmatism is an example of an internalist account that does this. Given that there are these ways of understanding foundationalism that violate the same constraints as Jeffrey conditioning, one might object that we cannot rule out Jeffrey conditioning as a form of foundationalism on this basis.

But notice that although the previous accounts allow that an agent's current foundational state can be changed later on, each is able to provide *some* feature that their foundations must have now. Like classical foundationalists, non-classical foundationalists assume that there is some property that picks out a foundational state—whether this is being a state that was formed by an unconditionally reliable process, or being a state that is an undefeated seeming, etc. By contrast, Jeffrey conditioning makes no such assumption. On the Jeffrey framework, evidence has no such distinguishing feature.

We've assumed that Jeffrey conditioning fails to be a form of foundationalism in virtue of failing to satisfy the first and third conditions identified above. However, there's a different way of understanding how Jeffrey conditioning fails to be a form of foundationalism, one that understands this failure as a violation of condition two.[14] Earlier we said that if there is no constraint on the partition that we update on—if there is no norm that tells us what this partition, or the values it gets assigned, ought to look like—then our account of updating is no stronger than Probabilism, or, equivalently, the Descriptive Account. And this is precisely what Jeffrey conditioning does. It places no constraints on the weights that our partition gets assigned. However, a different way of satisfying the Normative Approach would be to "go deeper" and take our basic state to be the *experience* that gives rise to our weighted partition. We could then take Bayesianism's fundamental updating constraint to be that we revise our beliefs in the way this experience prescribes.

But reconceiving the structure of Bayesian updating in this way isn't going to help us. While the Bayesian formalism includes a rule that regulates how a weighted partition gives rise to an update (recall, our evidence is the sufficient partition for our update), it does not include a rule that regulates how a phenomenal experience gives rise to an update. Since phenomenal experiences lack the inferential relation to updates that a basic state bears to non-basic states, they violate the second constraint on a foundationalist account identified above.

There's perhaps a way of retaining what is attractive about the previous proposal while avoiding what makes it problematic. It's been argued by McGrew (2010), among others, that one can overcome the problem I've just described by "propositionalizing" our experiences. Instead of assuming that experience must figure into relations with our priors, to deliver a posterior credence distribution, we might assume that each experience is associated with a proposition like, "John is having an experience that looks like a red sock", or something like this, that the agent gets with a credence of one. Then, the thought goes, we could simply update by Bayesian conditioning on this proposition.[15]

But while this solution resolves part of the previous worry by making our input the kind of thing that can stand in some inferential relation with our posterior credences, it

[14] The following line of argument has a steady, if diffused, presence in the literature on Jeffrey conditioning. There are references to it as early as Carnap (1957) (reprinted in Jeffrey (1975)) and as late as Weisberg (2009).

[15] Thanks to an anonymous referee for drawing my attention to this proposal.

doesn't resolve all of the previous worry. It's true that if we accept this proposal there will be no problem about how to incorporate non-propositional information into a framework designed for propositional evidence. But the proposal still leaves an important normative gap to be filled insofar as it fails to tell us exactly *which* proposition is justified with a credence of one. In the previous example, we might wonder why we ought to update on the proposition that "John is having an experience that looks like a red sock", rather than on the proposition that "John is having an experience that looks like a red-ish blue-ish sock", or the proposition that "John is having an experience that looks like a red-ish blue-ish green-ish sock". In order to capture the uncertainty and complexity of our experience, the proposition that we update on must be much more nuanced than the first proposition in this example. But even the last proposition in this example seems insufficient to do this justice. The propositionalizing strategy, then, leaves us with two worries. The first worry is that there may be no principled answer to the question of what minimal level of precision is needed to do justice to experience, with all of its subtleties. The second worry is that our language may be insufficiently precise to capture whatever this minimal level of precision turns out to be (*cf.* Christensen (1992, p. 543)). Perhaps these problems could be overcome. It would take a much longer discussion to establish that they could not. As things stand, however, I think these problems give us reason to pursue a different type of strategy.

If experiences are not better candidates for the role we are looking to fill, Jeffrey conditioning turns out to be strictly weaker than Bayesian conditioning since the latter includes a formal constraint on evidence that the former lacks. We are now also in a position to see why the Orthodox Account of diachronic coherence considered at the beginning of this discussion cannot help us to unify updates on certain and uncertain evidence. Recall this account says that Bayesianism's fundamental updating rule does not take a stand on what our evidence should look like, but tells us instead what we ought to do *supposing* we have some of this evidence in hand. The problem is that, even if we assume this interpretation of the Bayesian formalism, an update on certain evidence will entail certain other restrictions on our probability distribution in the future that Jeffrey conditioning does not entail. In particular, it will entail that we can no longer change our credence in any proposition that is inconsistent with our evidence. Adopting the Orthodox Account, then, will not allow us to avoid the strange consequence that updates on certain and uncertain evidence seem to pattern in very different ways.

Finally, one might object that if Jeffrey conditioning fits uneasily within Bayesian epistemology because it lacks a constraint on evidence, then so much the worse for Jeffrey conditioning. After all, most discussions treat Jeffrey conditioning as almost literally a footnote to the Bayesian program. But this objection misunderstands why discussions about Bayesian updating tend not to focus on Jeffrey conditioning. We idealize away from the cases that Jeffrey conditioning covers, not because they aren't important, but because we assume they will be covered in the same way as regular Bayesian conditioning. To show that they are not covered in this way is to undermine the Bayesian program in its entirety.

There's much to be said, then, for thinking of Bayesian conditioning as a form of foundationalism and for thinking that Jeffrey conditioning cannot be ascribed this same structure, and for further thinking that all of this is a problem. While I don't think it's impossible to adopt assumptions that would make Jeffrey conditioning a form of foundationalism, or assumptions that would yield the result that Bayesian conditioning does not actually have a foundationalist structure, there's a strong presumption in favor of not thinking of these frameworks in these ways. We should, then, be open to a different story about how to bring these two frameworks together.

5 Bayesian Coherentism

5.1 Commutativity as Coherence

Since Bayesian conditioning entails a constraint that makes it stronger than Jeffrey conditioning, putting these two updating rules on a par will require making Jeffrey conditioning stronger. But we've just seen that we can't make Jeffrey conditioning stronger by making it a form of foundationalism. Putting these two updating rules on a par, then, will require reinterpreting the norm that supervenes on the formal property that makes Bayesian conditioning so strong. In this section, I'll argue that we can do this by reconceiving Bayesian updating as a form of coherentism.

The fundamental difference between regular Bayesian conditioning and Jeffrey conditioning has always been assumed to be that the latter generalizes the certainty of evidence. A second notable difference in these frameworks is that only the regular framework is commutative over weighted evidence partitions. Only the regular framework makes the order in which we get evidence irrelevant to the credence distribution we end up with each and every time that we update. Many have argued that the fact that the Jeffrey framework is non-commutative is a significant mark against it.[16] Say I come to learn that my nephew's baseball team won their game. And then I come to learn that my boss has given me a raise. It does not seem as though receiving these pieces of evidence in reverse order should change the credences I end up with. Consistency seems to require that identical pieces of information be treated the same, no matter the order in which they are received. If I have the same information in two sequences of updates, a mere difference in the ordering of this information should not make a difference to anything else. (One might object that if we conceive of 'information' as something other than a weighted partition, we will not necessarily get the bad result that the Jeffrey framework is non-commutative. We'll consider this possibility more carefully in §5.4.)

While much discussed in the literature, the non-commutativity of Jeffrey conditioning has never been assumed to be a defining feature of it in the way that the uncertainty of evidence has been so understood. Instead it has been assumed to be an unfortunate, but non-essential, property of the Jeffrey framework. The Jeffrey framework has the property of sometimes yielding updates that aren't commutative over weighted evidence partitions. This suggests an intriguing possibility: Why *not* take the fundamental norm that grounds all Bayesian updates, including Jeffrey updates, to be that they minimize the defect of failing to commute. Why not take the norm for evidence that governs all updates

[16] For two early proponents of this view, see Domotor (1980) and Doring (1999).

to be, not that these updates be grounded in a basic state, as Bayesian foundationalism would have it, but that they be minimally non-commutative. This would mean understanding the formal norm for evidence that governs updates to be the requirement that the values these updates yield be as insensitive as possible to the order in which these updates were made. It would mean evaluating updates based upon the extent to which they are consistent in this way.

Whether or not this proposal for grounding Bayesian updates is reasonable depends upon whether we think that minimizing the extent to which updates fail to commute is a norm that Bayesians ought to be interested in. Given that so much has been made of the commutative property in the Bayesian literature, it's clear that it *is* a norm that Bayesians ought to be interested in. And I think we can say something even stronger than this; I think we can give the commutative norm an interesting gloss. I think that a norm that requires that we minimize the extent to which Bayesian updates fail to commute makes the Bayesian framework look like a form of coherentism about updating. In order to see this, it will, again, be useful to remind ourselves what coherentism looks like as a structure of justification that applies to beliefs.

Like traditional foundationalism, traditional coherentism assumes that the target of justification is a set of cognitive states. Unlike traditional foundationalism, traditional coherentism assumes that the locus of justification is not some particular cognitive state, but is, instead, the relations that some such states, our beliefs, stand in with one another. More specifically, traditional coherentism assumes that some set of beliefs is justified exactly when its component beliefs fit correctly, or cohere, with one another. A standard coherentist account includes measures of probabilistic coherence, logical coherence, and evidential coherence.[17]

Do these measures translate to the Bayesian updating framework? Logical coherence and probabilistic coherence are both preserved by Bayesianism's synchronic constraint. They are preserved no matter how we understand the structure of diachronic coherence for Bayesians. The interesting question, then, is what an account of evidential coherence will amount to in a Bayesian updating setting. What could it mean for a set of weighted evidence partitions to be consistent over updates?

It's well-understood what evidential coherence, or consistency, amounts to in the traditional setting. It's a measure of the degree to which some proposition confirms each other belief in the set to which it belongs. It's a measure of the degree to which every proposition in a set is *evidence for* every other proposition in that set. If I hold the belief that it will rain in a few hours (P_1), and also the belief that the owner of the shop down the street just put out her umbrella stand (P_2), then the belief that the baseball game will be rained out this afternoon (P_3), if it increases the proportion and strength of the inferential connections between the beliefs in this set, increases the evidential coherence of this set of beliefs. Traditional coherentism makes justification at least partly a function

[17] Of course, there are many coherentist accounts, and not all of them endorse all three of these constraints. Ewing (1934), for instance, takes coherence to be a matter of logical coherence alone, while Lewis (1946) takes coherence to be a matter of evidential coherence. Notably, Bonjour (1985) takes coherence to be a matter of logical and probabilistic coherence, as well as a number of other requirements that might be held to fall under the heading of evidential coherence (see pp. 97-99 for the details).

of the degree to which some set of propositions is consistent.[18]

I think we are now in a position to understand what Bayesian coherentism might amount to. We can triangulate on an account of Bayesian coherentism from the descriptions of Bayesian foundationalism and traditional coherentism we already have. From our earlier description of Bayesian foundationalism, we borrow the idea that what we are interested in are weighted evidence partitions or, equivalently, the updates they correspond to:

$$f_{\mathrm{E}}\colon \mathrm{UP} \mapsto \{0, 1\}$$

From the above description of traditional coherentism, we borrow the idea that the relation of justification we are interested in is consistency, where consistency is likely to come in degrees:

$$f\colon \{\mathrm{P}_1, \mathrm{P}_2, ..., \mathrm{P}_n\} \mapsto \mathbb{R}^+$$

Together, these commitments entail that the ideal of justification for the Bayesian coherentist is a consistent set of updates:

$$f\colon \{\mathrm{UP}_1, \mathrm{UP}_2, ..., \mathrm{UP}_n\} \mapsto \mathbb{R}^+$$

Since a notable way for an updating framework to treat its updates consistently is to require that they yield the same values whenever they happen, Bayesian coherentism is plausibly the requirement that our updates commute.

If all this is right, then an alternative to understanding Bayesian updating as a form of foundationalism is to understand it as a form of coherentism. We can state Bayesian foundationalism and Bayesian coherentism side by side to illustrate that each is a different way of filling out the Normative Approach:

> **Bayesian Foundationalism:** A probabilisitic belief transition is always such that,
>
> (a) there is a sufficient partition, $\{\mathrm{E_i}\}$, for this transition, and
>
> (b) an agent who undergoes this transition is diachronically coherent iff some $\mathrm{E_i}$ is held with certainty.
>
> **Bayesian Coherentism:** A pair of probabilisitic belief transitions is always such that,
>
> (a) there are sufficient partitions, $\{\mathrm{E_i}\}$, $\{\mathrm{F_j}\}$, for each of these transitions, and

[18] Many contemporary coherentist accounts spell this out probabilistically (for one notable account, see Fitelson (2003)). For instance, a very simple view might hold that, in the previous case, what accounts for the increased coherence provided by P_3 is that $p(P_1|P_2)<p(P_1|P_2 \wedge P_3)$ and $p(P_2|P_1)<p(P_2|P_1 \wedge P_3)$.

(b) an agent who undergoes these transitions is diachronically coherent to the extent that these transitions minimize the defect of failing to commute (what it means to minimize this defect will become clear in §6).

As I noted earlier, an interesting feature of Bayesian coherentism is that unlike either Bayesian conditioning (i.e., Bayesian foundationalism) or Jeffrey conditioning, it is undefined for a single update. Therefore it is not entailed by either Bayesian conditioning or Jeffrey conditioning. While this makes Bayesian coherentism an amendment to the traditional Bayesian framework, it is not an amendment that requires the Bayesian to take on any additional substantive commitments. No matter what other commitments one happens to hold, inconsistency will always be a prima facie defect. This explains the importance that Bayesians, and formal epistemologists in general, have placed on the commutative property. In effect, Bayesian coherentism represents a particular way of articulating a commitment that Bayesianism, as well as every other normative theory, already holds.

5.2 *Is Bayesian Coherentism a Coherence Constraint?*

We've just seen that Bayesian coherentism comprises two constraints: 1) a sufficiency condition, which gives us our partitions, and 2) a commutative constraint on these partitions and their weights. Even if we assume that the commutative constraint is a coherence constraint, one might argue that the sufficiency condition is best understood as a foundationalist constraint. A defining feature of the sufficiency condition is its "directed" nature. The change of values along such a partition is conceptually or schematically prior to the effects this change propagates throughout the rest of the agent's probability distribution; this change cannot run in the other direction. Thus there's a sense in which any Bayesian update is both uni-directional and acyclic. This is the sense that is captured by our understanding of Bayes nets as directed acyclic graphs. By contrast, traditional coherentism is characterized, and distinguished from foundationalism, by a *lack* of both of these features. One might, then, wonder whether my norm, which is grounded in these foundationalist features, is really a coherence constraint.[19]

While I agree with all of the above, I still want to insist that the sufficiency condition is most meaningfully understood as a coherence constraint, so that it's coherence all the way down. It's true that the existence of a sufficient partition implies that a Bayesian update proceeds in a particular direction. But, as we also noted earlier, the sufficiency condition is satisfied anytime an agent remains probabilistically coherent over time. Insofar as the sufficiency condition is satisfied just in case Probabilism is satisfied, and insofar as Probabilism is clearly a coherence constraint, I think the sufficiency condition is also properly regarded as a coherence constraint. Put a little differently, if we assume, as we did in §2, that the sufficiency condition is a deflationary or descriptive account of diachronic coherence, insofar as it entails that there is nothing more to being diachronically coherent than to remaining probabilistically coherent over time, we ought to regard

[19] Thanks to an anonymous referee for raising this objection.

the foundationalism that the sufficiency condition encodes as a deflated or descriptive, or merely schematic, form of foundationalism, insofar as this foundationalism is also grounded in coherence.

One might still object that my account diverges too significantly from the sorts of accounts of coherence that are already in the literature to be relevantly related to them—this time, in virtue of the commutative constraint that, I have argued, makes my account distinctively coherentist. One might object that my commutative constraint looks very different from both probabilistic accounts of coherence and, also, from the sort of evidential coherence defended by traditional epistemologists. To see the latter point most clearly, it suffices to notice that it's obviously possible for a non-commutative update to conform to the traditional coherentist's constraint. We might have a set of beliefs that don't commute but that, together, increase the support that they lend to each other.[20]

However, while it's true that the sort of coherence I am defending is different from probabilistic coherence and from the sort of evidential coherence that traditional epistemologists have in mind, it's also true that the latter types of coherence constraints are quite different from each other. I take it that what unifies all coherence constraints is that they impose wide-scope requirements: they impose requirements on how our attitudes fit together without taking a stand on the direction of fit.[21] Just as the question of whether our evidence coheres, in the traditional way, is a matter of whether our combined evidence yields some valuable epistemic property (namely, confirmation), so too is the question of whether our evidence coheres, on the Bayesian framework, a matter of whether our combined evidence yields some valuable epistemic property (namely, commutativity). It's true that the function of evidence differs on each of these frameworks: in the former case, it is that which confirms our beliefs, in the latter case, it is that which triggers an update. Accordingly, what it means for evidence to cohere will also differ. But my aim isn't to apply traditional coherentism to a different set of circumstances. My aim is to identify a structural feature that is essential to traditional coherentism and develop it in a different direction.

Finally, it's important to be clear that what's doing the heavy lifting in my account is the commutative constraint itself rather than the interpretation of the norm this constraint gives rise to as a form of coherentism. I think I've provided good reason for calling Bayesian coherentism a coherence constraint, but if one is still not convinced, I don't think anything is lost. Regardless of what we call it, we still have a constraint that unifies Bayesian updates on certain and uncertain evidence, or so I will now argue.

5.3 *A Unified Account of Bayesian Updating*

The final piece of the puzzle is to see how adopting Bayesian coherentism helps us with the problem of being able to say that both updates on certain and uncertain evidence proceed from the same normative structure. For, at first glance, it looks as though our initial problem persists: it looks as though Jeffrey conditioning bears the same relation to

[20] Thanks to an anonymous referee for raising this objection.

[21] For a discussion of the distinction between wide- and narrow-scope requirements, see, for instance, Broome (1999) and Kolodny (2007).

Bayesian coherentism that it bears to Bayesian foundationalism. Updates on uncertain evidence fail to be updates on basic states. But they also sometimes fail to be updates that commute. So is appealing to a commutative norm really any better than appealing to a foundationalist norm?

There *is* a relevant difference between these two sorts of appeals. What makes Bayesian foundationalism problematic is that adopting it would mean having to say that every update on uncertain evidence, qua update on uncertain evidence, is incapable of making the agent diachronically coherent. Only updates on certain evidence have foundationalist properties. Therefore, only updates on certain evidence can make an agent diachronically coherent.

But Bayesian coherentism does not have this same feature. While only updates on certain evidence have foundationalist properties, *both* updates on certain *and* uncertain evidence are capable of commuting. If we are looking for a norm to unify these two types of updates, then, a norm that makes diachronic coherence a matter of updates commuting is capable of fulfilling this function. The fact that updates on uncertain evidence, qua updates on uncertain evidence, are capable of satisfying the norm to commute, suggests that the best interpretation of why some set of updates on uncertain evidence have failed to commute is that they have failed to conform to Bayesian coherentism. They have failed to conform to an updating norm that requires that updates commute. By contrast, the fact that updates on uncertain evidence, qua updates on uncertain evidence, *aren't* capable of satisfying the norm to be an update on a basic state means that the *only* explanation for why updates on uncertain evidence fail to proceed from a basic state is that the foundationalist norm we would need to render this verdict just isn't there.[22]

In short, the fact that updates on uncertain evidence can't conform to a norm formulated in terms of a basic state entails that such updates aren't governed by Bayesian foundationalism. It entails that there is no such norm. By contrast, the fact that updates on uncertain evidence *are* capable of conforming to Bayesian coherentism suggests that they *can* be governed by Bayesian coherentism. Bayesian coherentism is plausibly a norm for such updates.

And I think we can say something even stronger than this. I think we can say that, not only are all updates on uncertain evidence capable of satisfying Bayesian coherentism, but that all updates on uncertain evidence *do*, to some extent, satisfy Bayesian coherentism. Identifying commutativity with coherence makes it natural to want to give commutativity a degree-theoretic interpretation. This move, which will happen in §6, will allow us to say that all Bayesian updates are diachronically coherent to a degree.

5.4 An Objection

Before moving on, I want to address what might look like an important objection to Bayesian coherentism. We've assumed that some Jeffrey updates are non-commutative in a way that ought to be avoided. However, Lange (2000) famously argues that although the Jeffrey framework is non-commutative over weighted evidence partitions, this fea-

[22] This follows from standard deontic logic, which says that a norm can't require X if X is logically impossible.

ture of it does not make it defective. This is because the framework does commute the *experiences* that underwrite belief revisions. More precisely, it commutes the Bayes factors we might plausibly identify with these phenomenal experiences. If we take experiences, rather than weighted partitions, to be the inputs to the updating process, we get the result that the Jeffrey framework does indeed commute its inputs. If this is the case, then no commutative constraint on weighted evidence partitions is needed.

It's true that there's an understanding of the Jeffrey framework that makes my solution unnecessary and, so, unmotivated. But it does this at the cost of denying that the Bayesian updating framework should be taken to be an account of epistemic justification in the first place. We saw, in the last section, that a problem with taking phenomenal experiences to be our foundations is that the Bayesian framework does not entail any rational way of mapping these phenomenal experiences to our updates. But if experiences aren't normative enough for the foundationalist, they aren't normative enough for the coherentist either. The lack of any rational way of mapping phenomenal experiences to our updates, via our Bayes factors, precludes our taking experiences to be the elements that ought to commute.[23]

It's for this very reason that Field (1978, p. 364) claimed that an account of how experience figures into the updating process could not be conceived of as a rational account, but could only be conceived of as "the problem of giving a complete psychological theory for a Bayesian agent". This aim is in tension with our inconsistent triad, which assumes that we have reason to want to square the formal features of Bayesian conditioning and Jeffrey conditioning with their normative features. While it's true that we can avoid an inconsistency between these two sets of features by insisting that the latter do not exist, I assume that, for many, this would be to give up the game.

Since assuming that experiences are the inputs to our updating framework isn't a promising strategy, we will need to assume that weighted partitions are the inputs to our framework. A commutative constraint on the weighted partitions we update on is, therefore, desirable.

6 Degree-Theoretic Commutativity

In the last section, I proposed an interpretation of the Bayesian framework that allows us to say that updates on certain and uncertain evidence proceed from the same normative structure. This proposal assumes that we can assess the incoherence of sets of updates based upon the extent to which they instantiate what has long been deemed to be a bad-making feature of the Bayesian formalism. If one wants to reject the proposal, one must either deny that (1) commutativity is an important feature for an updating rule to guarantee, or that (2) the fact that commutativity is an important feature for an updating rule to guarantee does not mean that it is an important feature for particular sequences of updates to have. Absent an argument for one of these claims, I assume we have good reason to proceed with the question of how a norm that draws on this intuitive idea might be

[23] For a more detailed discussion of this point, see Cassell (2020).

developed.[24]

While it goes beyond this discussion to consider and defend all the possible ways of developing such a norm, I think we can nonetheless talk in general terms about what such a norm would have to involve. An obvious place to start is with a formal property that is famously identified with the commutative property. Diaconis and Zabell (1982, p. 825-826) show that the property of "Jeffrey independence" is both necessary and sufficient for the commutativity of Bayesian updates. Here's what this property amounts to:

> **Jeffrey Independence:**
>
> Let P be a probability function. And let $P_{\mathcal{E}}$ and $P_{\mathcal{F}}$ be the probability functions that result from updating P on the partitions $\mathcal{E}=\{E_i\}$ and $\mathcal{F}=\{F_j\}$, respectively. The partitions $\mathcal{E}$ and $\mathcal{F}$ are Jeffrey independent with respect to P, $\{p_i\}$ and $\{q_j\}$ if $P_{\mathcal{E}}(F_j)=P(F_j)$ and $P_{\mathcal{F}}(E_i)=P(E_i)$ holds for all i and j.
>
> Thus Jeffrey independence says that Jeffrey updating on $\mathcal{E}$ with probabilities p_i does not change the probabilities on $\mathcal{F}$ and similarly with $\mathcal{E}$ and $\mathcal{F}$ interchanged.

The most straightforward way of developing the idea that commutativity is a form of diachronic coherence is to require that a sequence of updates be Jeffrey independent. However, if we are looking to mimic the concept of incoherence that we are borrowing from traditional epistemology, a degree-theoretic account of Bayesian diachronic coherence seems more appropriate.

How do we get a degree-theoretic account of Bayesian diachronic coherence? Having identified complete coherence with commutativity, and so with Jeffrey independence, one approach would be to define *in*coherence as a measure of the degree of a violation of Jeffrey independence for a sequence of updates. In other words, we might take the agent's degree of incoherence to be a measure of how much the weighted partitions we have updated on *fail* to be independent of each other.

How do we measure the degree to which some set of weighted partitions fail to be independent of each other? There is undoubtedly more than one way that a plausible measure for such a violation could be devised. It would take a much longer discussion to assess the merits of the different measures that might be appropriately put to use here. The aim of this paper is merely to suggest that this might be a discussion worth having.

Assuming that some measure represents a plausible way of giving content to the idea of diachronic incoherence we are interested in, the next question to ask is how this measure should be put to use in a norm. Again, since the aim of this paper is merely to motivate Bayesian coherentism as an alternative to Bayesian foundationalism rather than to

[24] *Can* we reject one of these two claims? I've already suggested that (1) seems unimpeachable. What about (2)? Perhaps one might want to argue that the kind of defect the non-commutativity of Jeffrey conditioning represents is not a defect of particular updates, but is a sort of defect that inheres in the framework in general. However, it's difficult to imagine what it could mean for a framework to be defective in a way that doesn't directly derive from the particular updates that it produces.

work out what the best version of Bayesian coherentism might turn out to be, I won't offer any suggestions about this. Perhaps we would want our norm to govern only pairs of sequential updates. Or maybe we would want our norm to govern larger sets of updates taken pairwise. However we choose to go, it's clear that the norm we end up with should look like the following:

> **Bayesian Coherentism (Revised):** A pair of probabilistic belief transitions is always such that,
>
> (a) there are sufficient partitions, $\{E_i\}$, $\{F_j\}$, for each of these transitions, and
>
> (b) an agent who undergoes these transitions is diachronically incoherent to the extent that these transitions fail to commute or, equivalently, fail to be Jeffrey independent.

It's a common idea that there are degrees of probabilistic incoherence.[25] This discussion introduces the idea that there may also be degrees of diachronic incoherence that aren't reducible to the latter by defending a degree-theoretic account of diachronic coherence that, unlike the Descriptive Account, isn't reducible to Probabilism. On the account that I've called Bayesian coherentism, perfect normative diachronic coherence is the special case where the agent's updates commute.

7 Bayesian Coherentism in Action

I've argued that Bayesian coherentism offers a unified account of Bayesian updates. It would take a much longer discussion to assess whether the advantages of adopting this updating rule outweigh the costs. However, it's important to at least establish that Bayesian coherentism yields plausible verdicts. In this section, I argue that it does.

To start, consider what might seem like a worry for Bayesian coherentism. Say I update on some weighted partition. And then I have an experience that justifies my updating again, in a way that violates Bayesian coherentism. Here it might seem as though there is a conflict between what my experience justifies and what Bayesian coherentism tells me to do.[26]

However, we've already seen that the commitments that motivate Bayesian coherentism ensure that no conflict like the previous one could possibly arise. The assumption that experience could justify a weighted partition is one that we have already rejected. We cannot have a conflict between what Bayesian coherentism justifies and what experience justifies if experience does not justify anything at all. One way of thinking about Bayesian coherentism, then, is as a solution to the input problem. It is the account of evidence we are left with once we acknowledge that we cannot get an account of evidence that is based in experience.

[25] See Schervish, Seidenfeld, and Kadane (2000, 2002, 2003). For a more recent account, see Staffel (2015).

[26] Thanks to an anonymous referee for raising this concern.

There is perhaps a more general worry in the neighborhood. Perhaps the real worry is that Bayesian coherentism diverges too significantly from what we expect of an account of evidence to be an account of diachronic coherence. However, I think this worry is also unfounded. Though Bayesian coherentism is a formal account of evidence, it is still able to capture certain intuitions we have about evidence. In the last section, we saw that failures of Jeffrey independence are necessary and sufficient for failures of commutativity. Diaconis and Zabell (1982, p. 825) also offer a less well-known, though equivalent, mathematical route to conditions for commutativity by noting that we can equally say that two updates will commute just in case the *second* update does not change the values along the *first* partition that has been updated on, and similarly when the order of these partitions has been reversed.[27] A different way of interpreting Bayesian coherentism, then, is as the requirement that the sufficient partitions for a series of updates have values that cannot be changed over time by each other.

This interpretation of Bayesian coherentism allows it to capture an intuition about evidence that I think is shared by many, namely, that unlike our credences, our evidence is not the sort of thing that is up for revision. Bayesian coherentism captures this trait of factive, as well as many other, accounts of evidence in a purely formal way by entailing that, to the extent that the values over some partition have changed over time, this partition fails to be evidence, properly speaking.[28]

This feature of Bayesian coherentism also gives rise to a more illuminating way of understanding how it unifies updates on certain and uncertain evidence. Earlier we said that certain evidence is distinguished from uncertain evidence by the fact that the former, but not the latter, cannot be lost or changed. This is what made updates on the former type of evidence, but not the latter, foundationalist. On the interpretation of the Bayesian framework that I am suggesting, *sets* of uncertain evidence have this very same feature. For we've just seen that Bayesian coherentism entails that an agent is diachronically coherent—that she has evidence—to the extent that the values over a sufficient partition received earlier are not changed by evidence received later, and to the extent that this similarly holds where the order of these evidence partitions has been reversed. *Both* updates on certain and uncertain evidence proceed from a coherentist norm of the very same kind, then, one that entails that an agent is diachronically coherent to the extent that the set of information she has updated on is unchangeable by the particular members of this set.

Before closing, it's worth briefly considering how the verdicts of Bayesian coherentism compare with those of approaches to modeling commutativity failure. To do this,

[27] The necessity claim is easy to see. Since Jeffrey updating fixes the probabilities on the partition (i.e., $P_{\mathcal{E}\mathcal{F}}(\mathrm{F_j}) = q_\mathrm{j}$ and $P_{\mathcal{F}\mathcal{E}}(\mathrm{E_i}) = p_\mathrm{i}$), an update will be commutative only if $P_{\mathcal{E}\mathcal{F}}(\mathrm{E_i}) = p_\mathrm{i}$ and $P_{\mathcal{F}\mathcal{E}}(\mathrm{F_j}) = q_\mathrm{j}$, for all *i* and *j* (p.825). Diaconis and Zabell don't give a proof of the sufficiency claim, but note that it follows from Csiszár (1975, theorem 3.2).

[28] Of course, Jeffrey himself did not defend this account of evidence. However, our discussion assumes that there are different ways of interpreting the core commitments of Jeffrey conditioning, and of the Bayesian updating framework more generally, that have different consequences for how we should think about evidence. This assumption seems especially plausible in light of the fact that Jeffrey's own views about his updating rule seem to have evolved over time.

we might consider Hawthorne (2004)'s famous example of a doctor who receives two different sets of diagnostic tests for lung cancer, one of which tests a sample for cancer-like cells, the other of which uses an x-ray to test for a mass. Hawthorne shows that the doctor's final probability in the proposition that their patient has cancer will depend upon the order in which these tests are received, which is clearly the wrong result.

One option for dealing with these sorts of cases is suggested by McGrew (2014)'s account of "targeted updating". Targeted updating reverses the orthodox order of things by telling us to *first* decide which proposition we are interested in updating and to *then* find a partition that satisfies the sufficiency conditions for that update. In the above case, the target proposition might be, "The patient has cancer". One way to apply McGrew's proposal, while preserving the commutativity of updates, would be to model the situation as one where both the experience of getting the sample results and the experience of getting the x-ray results happen simultaneously. In this case, the agent's evidence partition is plausibly a partition on the conjunction of the propositions these experiences give rise to.[29]

My approach agrees with Hawthorne's assessment and, also, offers an additional way of interpreting what goes wrong in such cases—one that focuses on properties of the agent's evidence rather than on the proposition that she uses this evidence to update. However, my approach differs from McGrew's in not offering a recommendation about which weighted partition the agent should update on. As an evaluative norm, all that it is in a position to say is *why* the agent's update exhibits some degree of diachronic incoherence, according to the more stringent standards of the Normative approach. While this response might disappoint, I have suggested already that we have independent reason to think that this is all that any Bayesian updating norm is in a position to offer. My norm supports models like the one that McGrew proposes without taking on those commitments that make such models objectionable qua action-guiding constraints. Indeed, there is an interesting parallel between McGrew's proposal and my own. Like McGrew, my constraint treats evidence non-diachronically. My constraint governs a set of evidence the agent has accumulated over time simultaneously without regard for the way in which this evidence was received.

Bayesian coherentism stands in need of more of a defense than I have been able to give it here. Incoherence on the Bayesian framework is taken to be more or less synonymous with the propensity for an agent's credal state to fail to maximize some sort of utility. Dutch book arguments famously show that where an agent fails to be coherent, her credal state fails to maximize practical utility by sanctioning a series of bets that would lead to a sure loss of money.[30] More recent accuracy arguments have established that where an agent fails to be coherent, she fails to have a credal state that minimizes inaccuracy.[31] The Bayesian norms that prescribe against the sort of incoherence that these

[29] See Diaconis and Zabell (1982, §4) for a similar proposal.

[30] See de Finetti (1937, 1972) for a justification of Probabilism, Lewis (1999) and Teller (1973) for a justification of Bayesian updating, and Skyrms (1987) for a justification of Jeffrey conditioning.

[31] See Joyce (1998, 2009) for justifications of Probabilism, Greaves and Wallace (2006) for a justification of Bayesian updating, and Leitgeb and Pettigrew (2010a, 2010b) and Pettigrew (2016) for justifications of both Probabilism and Bayesian updating.

arguments target are taken to be justified by these evaluative facts.

Whether Bayesian coherentism can be given this sort of defense is a question for another day. This question belongs to a more general discussion about whether formal evidential norms can be vindicated by dutch book arguments and accuracy arguments. Again, the aim of this paper hasn't been to offer a comprehensive defense of Bayesian coherentism, but to argue that this is the structure the Bayesian formalism would need to have in order to unify updates on certain and uncertain evidence. This discussion has assumed, as I think we always do, that we can identify some features that we would want the Bayesian framework to have, and *then* go on to ask whether the norms these features give rise to promote some sort of value. This paper has carried out the first part of this project. If, in the end, it turns out that there is nothing to be said for Bayesian coherentism other than that it does unify Bayesian updates—if it turns out that this norm does not in fact minimize inaccuracy or make the agent dutch book invulnerable—it may be that there is not an enormous amount of pressure to minimize incoherence in the way that Bayesian coherentism advises.

8 Final Thoughts

I want to conclude by reconsidering the motivation for this discussion. We have been assuming throughout that Jeffrey conditioning and regular Bayesian conditioning ought to be brought together; that updates on certain and uncertain evidence should proceed from frameworks with the same normative structure. But maybe they shouldn't. Maybe one can provide a principled explanation for why they don't. Maybe like Field (1978, p. 365) claims when discussing his own interpretation of Jeffrey conditioning, we would want to say that Bayesian conditioning is too much of an idealization to ever be of any use. Or maybe like Lange (2000, p. 397), we would want to hold that the conditions under which we get uncertain evidence differ in relevant ways from those under which we get certain evidence. Lange argues that our background beliefs only make a difference to the probative value of an experience in cases where this experience gives rise to uncertain evidence. Though he does not elaborate on why he thinks this, he does note at one point that cases where we update on evidence to which we have assigned a credence of one are cases where our background beliefs fail to function as "extended sense organs" (p.400), in the way that they do when we assign our evidence some other value.

Despite the weird imagery, this does not seem like a crazy suggestion. For starters, it does seem as though some evidence propositions, though they might be triggered by experience, are not plausibly the sorts of claims that could be *justified* by experience. When I change my credence in the proposition that a difficult math proof is correct, this change may be accompanied by certain sensory experiences that are brought about by introspection. However, these experiences do not seem to be what justify these revisions, in the way that my belief that the sky is blue seems to be justified by an experience with a certain phenomenal character. If this is the case—and if it is also the case that updates on certain evidence are exactly those updates that don't seem to be justified by experience—then the agent's background beliefs shouldn't have a hand in determining what her experience justifies since, in these cases, the agent's experiences do not justify anything at all.

More generally, the previous passage raises a possibility that we have not yet considered, which is that the *type* of content to which we happen to be justified in assigning a credence of one might differ in some important way from the type of content to which we happen to be justified in assigning a lesser value. If there is indeed this qualitative difference between our certain and uncertain evidence, the unified account that we have been after in this paper may be inappropriate. For while it may be implausible that there is a sharp cut-off between certain and uncertain evidence, there may very well be a sharp cut-off between the different types of propositions this evidence corresponds to. Updates on certain evidence might, then, be special in the way that a verdict that precludes all reasonable doubt is special, or in the way that a carpool lane that reduces pollution to an acceptable level is special. If updates on certain evidence are indeed special in this way, there may be reason after all to think that updates on certain and uncertain evidence should proceed from frameworks with different normative structures.

Of course, those who tell this sort of story owe us an account of why we might be justified in assigning only some particular class of propositions a credence of one. This might be a considerable task. Or it might not be. A modest proposal along these lines would be to appeal to the principle of Continuing Regularity. This principle says that we should assign probability one only to logical truths and zero only to contradictions (or, to necessary and impossible propositions, respectively). While not universally endorsed, this principle is believed by many Bayesians to be quite plausible. And since there is more or less agreement about which propositions are necessary and impossible, we would easily be able to identify the sorts of propositions to which we are justified in assigning a credence of one.

Maybe, then, there's some argument from Continuing Regularity to the conclusion that certain and uncertain evidence ought to proceed from frameworks with different normative structures. It would be interesting if there were. I think it's of interest to consider all of the possible ways that the fundamental commitments of the Bayesian framework can be articulated. One such way, which we have been considering here, is suggested by what appears as a footnote in nearly every paper on Bayesian updating. This is the assumption that Bayesian conditioning and Jeffrey conditioning are perfect parallels with respect to their formal structures. Since Bayesian updating is a normative theory, I have argued that it makes some sense to ask what it would mean for these updating rules to also be perfect parallels with respect to their normative structures. This paper has tried to answer this question. Maybe there's not much going for my answer besides its connection to the apparent truism that appears in all of these footnotes. But I do think it's interesting—and, also, surprising—to discover what this apparent truism ends up committing us to.

References

[1] Blackwell, David A and Girshick, Meyer A. 1979. *Theory of Games and Statistical Decisions.* Courier Corporation.

[2] BonJour, Laurence. 1985. *The Structure of Empirical Knowledge.* Harvard University Press.

[3] Bradley, Richard. 2005. "Radical Probabilism and Bayesian Conditioning." *Philosophy of Science*, 72(2): 342-364.

[4] Broome, John. 1999. "Normative Requirements." *Ratio*, 12(4): 398-419.

[5] Cassell, Lisa. 2020. "Commutativity, Normativity and Holism: Lange Revisited." *Canadian Journal of Philosophy*, 50 (2): 159-173.

[6] Csiszar. I. 1975. "Divergence Geometry of Probability Distributions and Minimization Problems." *The Annals of Probability*, 3(1): 146-158.

[7] Chrisman, Matthew. 2008. "Ought to Believe." *Journal of Philosophy*, 105(7): 346-370.

[8] Christensen, David. 1992. "Confirmation Holism and Bayesian Epistemology." *Philosophy of Science*, 59(4): 540-57.

[9] de Finetti, Bruno. 1937. "Foresight: Its Logical Laws, Its Subjective Sources", in Henry E. Kyburg and Howard E.K Smokler (eds.), Studies in Subjective Probability, Huntington, NY: Robert E. Kreiger Publishing Co.

[10] de Finetti, Bruno. 1972. *Probability, Induction and Statistics*. New York: Wiley.

[11] Diaconis, Persi and Sandy Zabell. 1982. "Updating Subjective Probability". *Journal of the American Statistical Association* 77.380: 822-830.

[12] Domotor, Zoltan. 1980. "Probability Kinematics and Representation of Belief Change". *Philosophy of Science*, 47: 384-403.

[13] Doring, Frank. 1999. "Why Bayesian Psychology is Incomplete". *Philosophy of Science* (Proceedings), 66: 379-389.

[14] Earman, John. 1992. *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. MIT Press.

[15] Ewing, A. C. 1934. *Idealism: A Critical Survey.* London: Methuen.

[16] Field, Hartry. 1978. "A Note on Jeffrey Conditionalization." *Philosophy of Science*, 45(3): 361-367.

[17] Fitelson, Branden. 2003. "A Probabilistic Measure of Coherence." *Analysis*, 63: 194-199.

[18] Goldman, Alvin. 1988. "Strong and Weak Justification", in *Philosophical Perspectives 2: Epistemology*, James Tomberlin (ed.), Atascadero, Calif.: Ridgeview Publishing Co., pp. 51-69.

[19] Harman, Gilbert. 1986. *Change in View: Principles of Reasoning.* MIT Press.

[20] Greaves, Hilary and David Wallace. 2006. "Justifying conditionalization: Conditionalization maximizes expected epistemic utility." *Mind* 115 (459):607-632.

[21] Hawthorne, James. 2004. "Three Models of Sequential Belief Updating on Uncertain Evidence." *Journal of Philosophical Logic,* 33 (1): 89-123.

[22] Howson and Urbach. 1989. *Scientific Reasoning: The Bayesian Approach*. Open Court Publishing Co.

[23] Huemer, Michael, 2007. "Compassionate Phenomenal Conservativism." *Philosophy and Phenomenological Research* 74 (1): 30-55.

[24] Jeffrey, Richard. 1965. *The Logic of Decision*. Chicago: Chicago University Press.

[25] Jeffrey, Richard. 1983. "Bayesianism with a human face", in John Earman (ed.), *Testing Scientific Theories,* 133-56. University of Minnesota Press.

[26] Jeffrey, Richard. 1975. "Carnap's Empiricism", in G. Maxwell and R.M. Anderson, Jr., (eds.), *Minnesota Studies in the Philosophy of Science*. Vol. 6, *Induction, Probability, and Confirmation*. Minneapolis: University of Minnesota Press, pp. 37-49.

[27] Joyce, James M. 1998. "A Nonpragmatic Vindication of Probabilism." *Philosophy of Science* 65 (4):575-603.

[28] Joyce, James. 2009. "Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In Franz Huber & Christoph Schmidt-Petri (eds.)" *Degrees of Belief.* Synthese. pp. 263-297.

[29] Kolodny, Niko. 2007. "How Does Coherence Matter?" *Proceedings of the Aristotelian Society (Hardback),* 107(1): 229-263.

[30] Lange, Marc. 2000. "Is Jeffrey Conditioning Defective in Virtue of Being Non-Commutative." *Synthese*: 123(3): 393-403.

[31] Leitgeb, Hannes and Richard Pettrigrew. 2010a. "An Objective Justification of Bayesianism I: Measuring Inaccuracy." *Philosophy of Science* 77 (2):201-235.

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

[33] Lewis, C.I. 1946. *An Analysis of Knowledge and Valuation.* Open Court.

[34] Lewis, David. 1999. *Papers in Metaphysics and Epistemology: Volume 2*. Cambridge University Press.

[35] McGrew, Lydia. 2010. "Probability Kinematics and Probability Dynamics." *Journal of Philosophical Research*, 35: 89-105.

[36] McGrew, Lydia. 2014. "Jeffrey Conditioning, Rigidity and the Defeasible Red Jelly Bean." *Philosophical Studies*, 168 (2): 569-582.

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

[38] Plantinga, Alvin. 1991. "Warrant and Designing Agents: A Reply to James Taylor." *Philosophical Studies* 64 (2):203 - 215.

[39] Pryor, James. 2000. "The Skeptic and the Dogmatist," *Nous*, 34 (4): 517-549.

[40] Schervish, Mark J and Seidenfeld, Teddy and Kadane, Joseph B. 2000. "How Sets of Coherent Probabilities May Serve as Models for Degrees of Incoherence". *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems* 8: 347-355.

[41] Schervish, Mark J and Seidenfeld, Teddy and Kadane, Joseph B. 2002. "Measuring Incoherence", *Sankhya: The Indian Journal of Statistics*, Vol. 64: 561-587.

[42] Schervish, Mark J and Seidenfeld, Teddy and Kadane, Joseph B. 2003. "Measures of incoherence: How not to gamble if you must." *Bayesian statistics* 7: 385-402.

[43] Skyrms, Brian. 1987. "Dynamic Coherence and Probability Kinematics." *Philosophy of Science* 54 (1):1-20.

[44] Skyrms, Brian. 1997. "The Structure of Radical Probabilism." *Erkenntnis*, 45: 285-97.

[45] Staffel, Julia. 2015. "Measuring the Overall Incoherence of Credence Functions." *Synthese*, 192 (5): 1467-1493.

[46] Teller, Paul. 1973. "Conditionalization and Observation." *Synthese*, 26: 218-258.

[47] Titelbaum, Michael. forthcoming. "Normative Modeling". *Methods in Analytic Philosophy: A Contemporary Reader.* J. Horvath ed. Bloomsbury Academic Press.

[48] Weisberg, Jonathan. 2009. "Commutativity or Holism? A Dilemma for Conditionalizers." *British Journal for the Philosophy of Science*, 60(4): 793-812.

[49] Weisberg, Jonathan. 2015. "Updating, Undermining and Independence." *British Journal for the Philosophy of Science*, 66(1): 121-159.