Bayesian Coherentism†

L. Cassell
lcassell@philos.umass.edu

Abstract. This paper considers a problem for Bayesian epistemology and goes on to propose 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 coherentism.

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 goes on to propose 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

†Reading Instructions: My presentation will focus mainly on pgs. 1-17.

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.

I will argue for an interpretation of the Bayesian framework that resolves the inconsistency by rejecting (3). I'll reject (3) by arguing that both regular Bayesian updates and Jeffrey updates proceed from frameworks with a coherentist structure.[1] 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 as a form of coherentism.

Here's how we'll get to this conclusion. In §1, I give some background. In §2, I describe the sense in which regular Bayesian updating has a foundationalist structure. In §3, I explain why adopting Jeffrey conditioning entails abandoning this foundationalism. In §4, I propose a constraint that looks like a version of coherentism about updating and argue that it more clearly supports the truism that Bayesian conditioning is a special case of Jeffrey conditioning. In §5, I give this constraint a formal backbone. Finally, in §6, I revisit the motivation for this

[1] In certain places, I will speak loosely and refer to updates on uncertain evidence as 'Jeffrey updates', or updates by Jeffrey conditioning. Strictly speaking, this is not correct, of course, since updates on certain evidence are also Jeffrey updates. However, in some contexts, it would be awkward to talk in any other way.

constraint.

1 Some Background

1.1 Diachronic Coherence for Bayesians

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, the agent's credences be a probability function. To say that a Bayesian agent is *synchronically coherent*, then, is to say that she conforms to Probabilism: 1) she assigns every proposition her credence function is defined over a non-negative value, 2) she assigns a credence of one to any tautology and, 3) for any mutually exclusive propositions, A and B, that her credence function is defined over, $cr(\mathrm{A}) + cr(\mathrm{B}) = cr(\mathrm{A} \vee \mathrm{B})$.

Second, Bayesianism is committed to the constraint that the agent's beliefs evolve over time in accordance with her conditional probabilities. If my credence in the proposition that I will play baseball tomorrow is .3, and my credence in the proposition that I will play baseball tomorrow *conditional* on the proposition that it will not rain is .7, then when I learn that it will not rain—when I get this as evidence—my credence that I will play baseball tomorrow should shoot up from .3 to .7. My current credence in any proposition ($p'(\mathrm{A})$) should always be my prior credence in that proposition, conditional on the evidence that I've gotten ($p(\mathrm{A}|\mathrm{E})$).[2]

To say that a Bayesian agent is *diachronically coherent*, then, is to say that her conditional probabilities guide her belief revisions. One way of capturing this idea is to require that the agent's current probabilities be determined by her conditional probabilities, in the way that we've just described. A different, though equivalent, way of capturing this idea is to require that the values of the conditional probabilities that yield the agent's current probabilities be the same before and after the update. 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.[3]

As it happens, every probabilistic belief transition has a set of arrows. For

[2] Where $p(\mathrm{A}|\mathrm{B}) = \frac{p(\mathrm{A} \wedge \mathrm{B})}{p(\mathrm{B})}$.

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

any probabilistic belief transition, there will be *some* information—like learning that it will not rain—that each of my beliefs are conditional on in the same way before and after this transition. More formally, there will always be a partition (a set of mutually exclusive and exhaustive propositions, like $\{\text{RAIN}, \overline{\text{RAIN}}\}$) that is sufficiently fine-grained to represent this transition as an update that is conditional on that partition:[4]

> Descriptive Diachronic Coherence for Bayesians: There is a sufficient partition for every probabilistic belief transition.

Or, equivalently, where $S=\{B_1, ..., B_n\}$ is a set of beliefs that form a partition, 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', at t and t', respectively, can be formulated in a way that underlines that there will always be some conditional probability that remains the same before and after the update, so that it can be understood to guide her belief revision:

> Descriptive Diachronic Coherence for Bayesians:
> $\forall p, \forall p', \exists S(\forall B_i \in S), \forall A(p(A|B_i) = p'(A|B_i))$, if defined.

Since Descriptive Diachronic Coherence for Bayesians is no stronger than Probabilism, one might suspect that it will be too weak to capture any interesting notion of diachronic coherence. To see that this is indeed the case, consider the very simple agent who only has beliefs about whether or not she will play baseball tomorrow. Her credences are only defined over the partition $\{\text{PLAY}, \overline{\text{PLAY}}\}$. Suppose these credences are $p(\text{PLAY})=.4$ and $p(\overline{\text{PLAY}})=.6$. Suppose further that the agent revises her beliefs to $p(\text{PLAY})=.7$ and $p(\overline{\text{PLAY}})=.3$. In this case, there is a partition that is sufficient for the update: $\{\text{PLAY}, \overline{\text{PLAY}}\}$. Therefore, this belief transition satisfies Descriptive Diachronic Coherence for Bayesians. Intuitively, however, this case doesn't look much like an agent responding to her evidence. For we tend to think that updating in accordance with one's evidence happens when we come to change our belief in some proposition, on the basis of some *different* information. It happens, as in the example above, when our views about whether we will play baseball tomorrow change in response to listening to the weather forecast and learning about the chance of rain. The lesson is that some belief transitions that satisfy Descriptive Diachronic Coherence for Bayesians don't look much like an agent being

[4] For the proof of this, see 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.

diachronically coherent at all, if we take such coherence to involve the agent getting evidence. Instead, what they look like is an agent swapping one set of probabilities for another.

We can remedy this by strengthening our account of diachronic coherence. We can do this by stipulating that it is only when the agent conditions her beliefs on partitions that meet some additional constraint for being evidence that she is diachronically coherent:

> Normative Diachronic Coherence for Bayesians: A probabilisitic belief transition ought to be such that:
>
> (a) there is a sufficient partition, $\{B_i\}$, for the transition, and
>
> (b) $\{B_i\}$ satisfies the conditions for being evidence.

Or, equivalently, where $E = \{B_1, ..., B_n\}$ meets the criteria for being evidence, the following prescribes the relation between an agent's prior credence distribution p and her posterior credence distribution, p', at t and t', respectively, by means of the obligatory operator, O:

> Normative Diachronic Coherence for Bayesians:
> $\forall p, \forall p', \forall B_i \in E, \forall A, O(p(A \mid B_i) = p'(A \mid B_i))$, if defined.

What makes this formulation normative is that it is stronger than Probabilism: an agent might transition from one probability function to another in a way that violates it. What makes this formulation a norm of coherence is that it is defined over a set of credence functions. Finally, what makes this norm of coherence diachronic is that these credence functions are indexed to different times.

Before moving on, I want to mention one last way of understanding diachronic coherence for Bayesians: a middle path between descriptive and normative diachronic coherence. Instead of overcoming the weaknesses of the former by restricting the conditions under which some partition is evidence, we might simply take for granted the existence of an evidence partition, and ask about what follows from it. In other words, we might take the Bayesian agent's diachronic obligations to consist in how she ought to proceed, *assuming* that she has a certain piece of evidence. On this picture of things, the evidence partition is not normatively determined, but *causally* determined: it is "an internal or psychological condition that must be checked or accepted at each stage."[5]

This understanding of diachronic coherence looks like a more modest way of getting us what we are after. By stipulating that some partition of propositions constitutes the agent's evidence, it avoids the worry that it is too weak to

[5] Diaconis and Zabell (1982, p. 825).

capture any interesting notion of diachronic coherence. But since this understanding of diachronic coherence does not require an agent's evidence to satisfy any additional constraints, it is also no stronger than Descriptive Diachronic Coherence for Bayesians. Moreover, it makes sense of the way that people tend to talk about Bayesian updating. Some might even call this *the default view* of diachronic coherence for Bayesians.

I want to defer saying anything more about the default view 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.

1.2 A Formal, Deflationary Account of Evidence

For now, then, let us assume that the aim of this paper will require a normative account of 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 such 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, then, is a constraint on evidence formulated in terms of a property that is not already part of the Bayesian formalism.

This paper will take a different approach by defending a *formal* account of evidence: a constraint on evidence that *is* formulated in terms of some feature of the Bayesian formalism. A Bayesian formalist about 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—whether or not these weighted partitions are further justified by the sorts of substantive considerations that we have just mentioned.[6] Exactly what it will look like for a formal

[6]The distinction between formal and substantive norms roughly tracks the distinction between 'thin' normative concepts like consistency—concepts that anyone, regardless of their other normative commitments would have to concede is a *prima facie*, good-making feature—and the sorts of thick normative concepts that are capable of distinguishing normative views (for the canonical account of the distinction between thin and thick normative concepts in the moral domain, see Williams (1985), especially pp. 140-142, 150-152.).

Perhaps an easier way of understanding what makes knowledge and reliability substantive, rather than formal, 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 treats evidence consistently will differ from the way that a defender of Dempter-Shafer theory treats evidence consistently. While the Bayesian will spell out her notion of consistency by means of probability functions, a Dempster-Shafer theorist, who trades in belief functions (or mass functions), will cash out her notion of consistency in terms of these. Unlike knowledge or reliability, constraint on evidence to be satisfied will become clearer in just a little bit. For now, 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.

The reason I defend a formal account of evidence in this paper is 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 emphasizing that a formal account of evidence will be a deflationary one. My proposal takes seriously the idea that there is nothing more to being a constraint on evidence than being a constraint on the sufficient partition of a belief transition. Insofar as we are tempted to talk of weighted partitions as "being evidence", then, it is because the weights these partitions get assigned are what determine the extent to which our constraint on evidence gets satisfied.[7] This deflationary picture of evidence will allow us to develop an account of normative diachronic coherence with the following features:

1. Agents aren't diachronically coherent, full-stop. Instead, they are diachronically coherent to varying degrees.

2. The constraint that determines an agent's degree of diachronic coherence isn't defined over weighted partitions. Instead, it's defined over *sets* of weighted partitions.[8]

3. The sets of weighted partitions that our constraint is defined over isn't assigned to an agent at a single time. Instead, it's assigned to temporally extended sequences of the agent.

We've said that an agent's standing as synchronically coherent will depend upon how her credence functions are related to each other. In defining diachronic coherence over sets of weighted sufficient partitions that the agent has at different times, my account entails that an agent's standing as diachronically coherent will depend upon how her *updates* are related to each other. My account, then, leaves us with an interpretation of Bayesian epistemology that is coherentist, with respect to both strands of justification that run through it.

---

the norm of consistency is so thin that it isn't complete, absent a framework that gives it content.

[7]Nevertheless, I will continue to talk in this way.

[8]A consequence of this is that NORMATIVE DIACHRONIC COHERENCE FOR BAYESIANS is more perspicuously formulated in terms of a pair of evidence partitions, instead of just one. How this can be done will become clearer in §4.

1.3 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, 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 way, has received its fare 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 (1988, 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 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.[9]

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.[10] This means that although Bayesian epistemology sets certain standards, there are no obligations issued by the theory. Just as we can say that cars are

[9] See also Kornblith (1992), 910, for a similar version of the objection.

[10] Defenders of evaluative norms in general include Feldman (2001) and Wolterstorff (2010). (Note that I will continue to use the word "norm" to refer to evaluative standards, even though most put the normative and the evaluative at odds with each other. I do so mainly for ease of exposition. But also because it seems intuitive (at least to me) that there might be norms for states of affairs, in addition to norms for agents. For discussion of this point, see, for instance, Chrisman (2008).)

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.

The way that we've set things up in this section already points us in the direction of conceiving 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, as we have done above, 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 obligate the agent to update in any particular way.

2 Bayesian Conditioning: Foundationalism about Updating

We've said that, without a constraint on evidence, the sort of diachronic coherence that Bayesian updating involves amounts to no more than Probabilism. But where one of the members of the sufficient partition of an update is a certainty, such an update *does* include a constraint on evidence: it includes a foundationalist constraint on evidence.[11] To see this, we will need to get clear on what foundationalism amounts to when applied to updates. And in order to do *this*, we will need to get clear on what foundationalism amounts to when applied 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: (1) that some cognitive states are basic, in the sense of their being justified not in virtue of their relations to other cognitive states and, (2) that all non-basic states are justified in virtue of some relation that they bear to basic states. In addition, classical foundationalism assumes (3) that the distinguishing mark of basic states is their infallibility.[12] Given this, one obvious candidate for an agent's basic state on the Bayesian framework is her evidence. We can formulate a constraint that captures this idea by focusing, once

[11]Those who have explicitly taken standard Bayesian conditioning to instantiate a foundationalist structure include Christensen (1992), Skyrms (1997), Bradley (2005), and Weisberg (2009), among others.

[12]By contrast, many recent, non-classical foundationalist accounts, like Goldman's (1988) reliabilism, Plantinga's (1991) proper basicality, Pryor's (2000) dogmatism and Huemer's (2001) phenomenal conservativism defend some form of fallible foundationalism. That is, they maintain that the property that makes beliefs basic is something other than their infallibility.

again, on the relation that Bayesian updating secures between an agent's conditional probabilities and the credence function they direct her to adopt:

> Bayesian conditioning: If the strongest evidence you get raises your credence in B to one, then your new degree of belief in A, for any A, should be $p'(\mathrm{A}) = p(\mathrm{A} | \mathrm{B})$, where A and B are propositions.

Where we assume the sort of Cartesian foundationalism that identifies infallibility with certainty, Bayesian conditioning satisfies (3) by requiring that some proposition in the agent's evidence partition receive a value of one—by requiring that it be a proposition of which she is certain. This formulation clearly satisfies (2) as well: the values we assign the rest of our beliefs depend upon our evidence. 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 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.[13] Therefore, 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. Perhaps most importantly of all then, (1) is satisfied as well.

Foundationalism, then, is a structure that applies just as easily to updates as it does to beliefs. Traditional foundationalism makes justification a function of whether some belief (P) is in the set of beliefs justified by an agent's basic state (S):

Traditional foundationalism: f_{S}: $\mathrm{P} \rightarrow \{0, 1\}$

(1 if $\mathrm{p} \in \mathrm{S}$, and 0 otherwise)

By contrast, where we take an update (UP) to be a probabilistic belief transition, Bayesian foundationalism says that Normative Diachronic Coherence for Bayesians is satisfied when (UP) is in the set of updates justified by the agent's evidence (E), where the constraint on evidence is the foundationalist constraint described by Bayesian conditioning:

Bayesian foundationalism: f_{E}: $\mathrm{UP} \rightarrow \{0, 1\}$

(1 if $\mathrm{up} \in \mathrm{E}$, and 0 otherwise)

There are a few things to notice about Bayesian foundationalism. First, unlike traditional foundationalism, Bayesian foundationalism governs an *update*. It tells us what our beliefs ought to look like in the future, rather than whether they

[13] That certainties stay certainties is simply a mathematical feature of the formalism.

are justified at any given moment. Second, Bayesian foundationalism governs the *values* we assign these beliefs. Finally, and most importantly, the constraint that Bayesian foundationalism imposes is a merely formal one. While it requires the agent's probability function to be encoded with certain values before and after an update, there is no further norm that underwrites the assignment of these values. These values constitute a form of foundationalism, regardless of whether or not they are justified by some further substantive consideration.[14]

3 Jeffrey Conditioning: Foundationalism Undermined

Most take the fundamental idea behind Jeffrey conditioning to be the thought that, as Jeffrey (1983, p. 171) himself 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.

In order to capture this more realistic class of cases, we need a rule that tells us how we ought to revise our beliefs whenever we get this sort of uncertain evidence. Jeffrey (1965) introduces a rule that does just this by allowing our evidence to assign values other than zero and one to the members of our partition:

> Jeffrey conditioning: If experience directly changes your credences over a partition $\{B_i\}$ from $p(B_i)$ to $p'(B_i)$, then your new degree of belief in A, for any A, should be $p'(A)=\sum_i p(A| B_i)p'(B_i)$.

It's clear from this formulation of it that 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, for instance, allowing us to assign probabilities other than zero and one 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, for instance, allowing us to assign probability one to the possibility that the sock is red and probability zero to the possibility that it isn't.

[14]One might object that we defined a formal constraint on evidence in §1.2, not as a constraint on the agent's entire credence function but, rather, as a constraint on the sufficient partition of her update. But, of course, since the values of a sufficient partition entail values for the credence distribution it is sufficient for, these amount to the same constraint.

But although Bayesian conditioning is a special case of Jeffrey conditioning, it is a degenerate case of it. This is because Jeffrey conditioning lacks the foundationalist constraint that governs Bayesian conditioning. There are a couple of ways of understanding how this structure is lacking. First, assume that we take the agent's basic state to be the evidence partition that she updates on. Since the propositions in this partition can receive a value of less than one—less than complete certainty—it doesn't include an infallible belief. More importantly, since the propositions in this evidence 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, the beliefs that comprise these evidence partitions violate the first and third conditions of foundationalism identified above.

Here's a different way of understanding how Jeffrey conditioning fails to be a form of foundationalism.[15] Earlier we said that if there is no constraint on the evidence partition that generates a particular belief transition—if it can generate any belief state that is consistent with Probabilism, by receiving any set of values consistent with Probabilism—then it fails to satisfy Normative Diachronic Coherence for Bayesians. Given this, it's tempting to think that if we want to satisfy Normative Diachronic Coherence for Bayesians, we should take the values that a partition gets assigned to be constrained by the *experience* that gives rise to it. That is, we should take this experience, rather than our evidence, to be our basic state. However, the Bayesian formalism does not regulate how experience gives rise to an update. Since experiences lack the inferential relation to updates that a basic state bears to non-basic states, they aren't better candidates for the role we are looking to fill. On this understanding of things, our updating rule violates the second condition of foundationalism identified above.[16]

Therefore, Jeffrey conditioning is strictly weaker than Bayesian conditioning: the latter includes a formal constraint on evidence that the former lacks. These considerations also show us why the default view of Bayesian diachronic coherence considered earlier can't help us. Recall this is the view that takes an agent's diachronic obligations to follow from an evidence partition that we have assumed the agent to have gotten. Since the default view is not committed to foundationalism, it is strictly weaker than regular Bayesian conditioning. Therefore, it will be too weak to serve as an account that can unify Bayesian updates.

How *do* we unify Bayesian updates then? Since Bayesian conditioning entails a constraint that makes it stronger than Jeffrey conditioning, putting these

[15]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)).

[16]Here, again, we note that to impose a constraint on an evidence partition just is to impose a constraint on an agent's credence distribution.

two updating rules on a par will require making Jeffrey conditioning stronger. But 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 normative consequences of the formal property that makes Bayesian conditioning so strong. It will require finding a norm capable of governing all Bayesian updates. The rest of this discussion will propose and defend a norm that does just this.

4 A Solution: Coherentism about Updating

We've just seen that Jeffrey conditioning lacks the formal constraint on evidence that makes regular Bayesian conditioning a form of foundationalism. In this section, I'll argue that we can get a unified account of Bayesian updating—one that allows us to claim that updates on certain and uncertain evidence proceed from frameworks with the same normative structure—by reconceiving of 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 evidence partitions: only the regular framework makes the order in which we get evidence irrelevant to the credence distribution we end up with at each and every time that we update. This is a significant mark against the Jeffrey framework. Consistency seems to require that identical pieces of information be treated the same, no matter the order in which they are received. And Jeffrey updates aren't guaranteed to be consistent in this way.

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 defect of the Jeffrey framework.[17] This suggests an intriguing possibility: why *not* take the fundamental norm that governs all Bayesian updates, including Jeffrey updates, to be that they minimize

[17] See Domotor (1980) and Doring (1999). For the classic rebuttal of the charge that Jeffrey conditioning is defective in virtue of being non-commutative, see Lange (2000). Lange claims that though the Jeffrey framework isn't commutative over evidence, this does not entail that it isn't commutative over the experiences that underwrite belief revisions. Therefore, it isn't non-commutative in a way that makes it defective.

In other work, I argue that Lange's argument does not target Jeffrey conditioning, but a more sophisticated updating rule. Moreover, I argue that it is not even an adequate defense of this updating rule. Therefore, there is reason to think that the Jeffrey framework is indeed defective, in virtue of not being commutative over evidence. Or, as I will argue in §6, it is defective, in virtue of not being commutative over evidence, provided that there are no other relevant normative considerations in play.

the defect of failing to commute. Why not take the norm for evidence that governs all updates to be, not that these updates be grounded in a certainty, 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 requiring that all updates be consistent in this way.

Whether or not this way of grounding Bayesian updates is a reasonable move to make 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. Moreover, I think we can give this 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. To get to the conclusion that Bayesianism is a form of coherentism about updating, it will again be useful to consider what this structure of justification looks like when it is applied to beliefs.

Like traditional foundationalism, traditional coherentism assumes that the locus of justification is a set of beliefs. It assumes that some set of beliefs is justified exactly when its component beliefs cohere, or fit correctly, with one another. On many coherentist accounts, probabilistic coherence, logical coherence, and evidential coherence are each measures that contribute to a belief set's coherence.[18]

Logical coherence and probabilistic coherence will both be preserved over time by Bayesianism's synchronic constraints: they will be 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 setting. It's well-understood what evidential coherence amounts to in a static setting. It is a measure of the degree to which some proposition confirms each other belief in the set to which it belongs. It is a measure of the degree to which every proposition in a set is *evidence for* every other proposition in the 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

[18]There are, of course, many different kinds of coherentist accounts of justification, both historically and contemporaneously. And many earlier coherentists did not 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 probabilistic coherence alone. 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).

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.[19] Traditional coherentism makes justification a function of the degree to which some set of propositions, $P_1, P_2, \ldots P_n$, cohere :

Traditional Evidential Coherence: f: $\{P_1, P_2, \ldots P_n\} \rightarrow \mathbb{R}^+$

We can triangulate on an account of Bayesian coherentism from the descriptions of Bayesian foundationalism and traditional coherentism we already have. From *traditional coherentism*, we borrow the idea that the locus of justification is a consistent set. From *Bayesian foundationalism*, we borrow the idea that the target of this justification are the values triggered by updates, rather than the contents of beliefs. What we are interested in isn't whether the contents of some set of beliefs are consistent at a time, but whether the values of some set of updates are consistent over time. The ideal of justification for the Bayesian coherentist is a consistent set of updates, UP, where, again, consistency comes in degrees:

Bayesian Evidential Coherence: f: $\{UP_1, UP_2, \ldots UP_n\} \rightarrow \mathbb{R}^+$

Therefore, where we take a set of updates to be a set of probabilistic belief transitions, the Bayesian coherentist will say that NORMATIVE DIACHRONIC COHERENCE FOR BAYESIANS is satisfied to the extent that one's updates are coherent, or consistent. The underlying requirement is that the evidence partitions implied by a set of updates be treated consistently. Since the most obvious way for an updating framework to treat consistently the evidence partitions implied by a set of updates is to require that they yield the same values whenever we get them, this version of NORMATIVE DIACHRONIC COHERENCE FOR BAYESIANS looks like the requirement that 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. I'll go on to say more in the next section about what Bayesian evidential coherence amounts to. In particular, I'll offer a proposal for how this sort of coherence can be represented as a gradable property. But, before we do that, it will be useful to get a feel for where we are right now. We can state Bayesian foundationalism and Bayesian coherentism in a way that illustrates that each gives us a different version of NORMATIVE DIACHRONIC COHERENCE FOR BAYESIANS:

BAYESIAN FOUNDATIONALISM: A probabilisitic belief transition will be such that:

[19]Contemporary coherentist accounts tend to spell out the notion of an inferential connection probabilistically. For instance, some have said that, in the previous example, what accounts for the increased coherence provided by P_3 is that $p(P_1 \mid P_2) < p(P_1 \mid P_2 \wedge P_3)$ and $p(P_2 \mid P_1) < p(P_2 \mid P_1 \wedge P_3)$.

(a) There is a sufficient partition, $\{E_i\}$, for the transition.

(b) It is diachronically coherent iff some E_i is held with certainty.

Bayesian coherentism: A pair of probabilisitic belief transitions will be such that:

(a) There are sufficient partitions, $\{E_i\}$, $\{F_j\}$, for each of these transitions, and

(b) They are diachronically coherent to the extent that they maximize evidential coherence (or minimize evidence incoherence), in a sense that will be made more precise in the following section.

As I alluded to 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 Jeffrey conditioning or Bayesian conditioning. Notice, however, that while this makes Bayesian coherentism an amendment to the traditional Bayesian framework, it is not an amendment that requires this framework to take on any additional substantive commitments. No matter what other commitments one maintains, 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, what Bayesian coherentism represents is just a different way of articulating a commitment that Bayesianism, as well as every other normative theory, already holds.

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 regular updates and Jeffrey updates proceed from frameworks with the same normative structure. For, at first glance, it looks as though this problem persists. It looks as though Jeffrey conditioning bears the same relation to Bayesian coherentism that it bears to Bayesian foundationalism. Jeffrey updates fail, in general, to be updates on basic states. But they also fail, in general, to be updates that commute. So, is appealing to a commutative norm really any different than appealing to a norm that makes justified belief revision a matter of maintaining a certain relation with some basic state?

I think 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 Jeffrey update, qua Jeffrey update, is incapable of making the agent diachronically coherent. Trivially, updates on uncertain evidence aren't capable of being updates on certain evidence. And updates on certain evidence are the only updates that have foundationalist properties.

But Bayesian coherentism would not have this same feature. This is because 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 Jeffrey updates fail to commute is that they have failed to conform to Bayesian coherentism. 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, entails that the foundationalist norm that we would need to render this verdict just isn't there.[20]

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 Jeffrey updates *are* capable of conforming to Bayesian coherentism suggests that they are governed by Bayesian coherentism. It suggests that Bayesian coherentism is 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* satisfy Bayesian coherentism—to some extent. I've already suggested that coherence is most plausibly interpreted as a gradable property. Identifying commutativity with coherence, then, makes it natural to want to give commutativity a degree-theoretic interpretation, as I do in the following section. This will enable us to say that all Bayesian updates are diachronically coherent, to a degree.

5 Bayesian Evidential Incoherence

In the last section, I proposed a way of grounding Bayesian updates that would allow us to say that updates on certain and uncertain evidence proceed from the same normative structure. This proposal rests in the intuitive idea 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, then 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 individual sequences of updates to have. Absent an argument for at least one of these claims, I assume that we have good reason to proceed with the question of how a norm

[20]This follows from standard deontic logic, which says that a norm can't require X if X is logically impossible. Thanks to Chris Meacham for helping me to clarify this point.

that draws on this intuitive idea might be developed.[21]

As I've noted already, many Jeffrey updates will commute. Therefore, the simplest way of developing the proposal of the previous section is to say that some sequence of updates satisfies NORMATIVE DIACHRONIC COHERENCE FOR BAYESIANS iff it commutes, and fails to satisfy NORMATIVE DIACHRONIC COHERENCE FOR BAYESIANS iff it fails to commute. We can spell out this idea in formal terms by appealing to a property that is both necessary and sufficient for commutativity. This is the property of Jeffrey independence:[22]

> JEFFREY INDEPENDENCE (JI): 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 $\{p_i\}$ 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 unifying certain and uncertain updates under a commutative norm, then, is to require that a sequence of updates be Jeffrey independent. However, if we are looking to mimic the concept of evidential coherence—or evidential *in*coherence—that we are borrowing from traditional epistemology, we will want a degree-theoretic account of this. How do we get a degree-theoretic account of evidential incoherence? If we identify complete diachronic coherence with Jeffrey independence, then an obvious approach to partial diachronic coherence is to quantify the degree of a violation of Jeffrey independence for a sequence of updates. In the appendix, I develop and defend a measure that does just this. Here's what this measure ends up looking like:

> EVIDENTIAL INCOHERENCE (EI): Let P be a probability function, and let $P_{\mathcal{E}}$ be the probability function that results from updating P on $\mathcal{E}$ with probabilities p_i. Finally, let $r_{ij}=\frac{P(E_iF_j)}{P(E_i)P(F_j)}$, $r_{ij'}=\frac{P_{\mathcal{E}}(E_iF_j)}{P_{\mathcal{E}}(E_i)P_{\mathcal{E}}(F_j)}$, if

[21]*Can* we reject one of these two claims? I've suggested that the first claim seems unimpeachable: the formal epistemology literature seems to care a lot about the commutativity of formal updating rules.

What about (2)? Perhaps one might want to argue that the kind of defect the non-commutativity of updates represents is not a defect of particular updates, but is a sort of inconsistency that inheres in the framework in general. However, it's difficult to imagine what it would mean for the framework in general to be defective, in a way that doesn't accrue to particular updates. I think, then, that we are safe in proceeding.

[22]The term 'Jeffrey independence' was coined by Diaconis and Zabell (1982).

defined, and 1 otherwise.[23]

> A sequence of updates over the partitions, $\{E_i\}$, $\{F_j\}$, is coherent to the extent that it minimizes $\sum_j(|1\text{-}\sum_i r_{*,j} p_i|)+\sum_i(|1\text{-}\sum_j r_{i,*'} q_j|)$.

It's an interesting question how EI might best be put to use in a norm. Since the aim of this paper is merely to establish Bayesian coherentism as an alternative to Bayesian foundationalism, I won't consider that question here. Perhaps we would want our norm to govern only pairs of updates that are sequential—that happen one after the other. Or maybe we would want our norm to govern larger sets of updates taken pairwise. But however we choose to go, it's clear that the kernel of the norm we would want is represented by the description of minimizing incoherence that EI encodes. This description allows us to sharpen our formulation of Bayesian coherentism in the following way:

> Bayesian coherentism (Revised): A pair of probabilistic belief transitions will be such that:
>
> (a) There are sufficient partitions, $\{E_i\}$, $\{F_j\}$, for each of these updates, and
> (b) Their degree of diachronic incoherence is determined by EI.

It's a common idea that there are degrees of probabilistic incoherence. This discussion introduces the idea that there are also degrees of diachronic incoherence that aren't reducible to the latter by defending a degree-theoretic account of Normative Diachronic Coherence for Bayesians that isn't reducible to Probabilism—that isn't reducible to Descriptive Diachronic Coherence for Bayesians. On the account that I've called Bayesian coherentism, perfect normative diachronic coherence will be the special case where the agent's updates commute.

6 Final Thoughts

Say I am told there's a thirty percent chance of rain tomorrow by Jack. And then I am told there's a seventy percent chance of rain tomorrow by Jill. Given certain

[23]One might worry about the 'if defined' clause in this formulation, which is meant to deal with those cases where we update on a certainty. It might be objected that this ad hoc fix undermines the aim of this paper, which is to unify updates on certain and uncertain evidence under the same norm. But requiring this clause in cases where an update on certain evidence makes this measure undefined is innocuous. However we choose to iron out the details our evaluative norm, it will always be the case that updates on certain evidence *are* consistent, in virtue of commuting. It is this intuitive notion of consistency that binds certain and uncertain updates, rather than the perhaps inelegant way that we are forced to give this notion formal content.

plausible assumptions, if these experiences cause me to revise my beliefs, they may very well yield two updates that don't commute. If I update twice on the proposition that it will rain tomorrow, my final credence that it will rain tomorrow will be .7. But had I gotten Jack and Jill's testimony in reverse order, my final credence that it will rain tomorrow would have been .3. Therefore, Bayesian coherentism will say that these updates are not perfectly diachronically coherent.

But now suppose that Jill is reliable, when it comes to matters of the weather, whereas Jack isn't. If my updates are the same as before—with the evidence that Jill's testimony gives rise to swamping the evidence that Jack's testimony gives rise to—Bayesian coherentism will again tell me that I'm not perfectly diachronically coherent. But is this still the right result?

The account developed in the previous two sections says that the formal constraint that guides all Bayesian updates is that they be made in a way that makes the order in which evidence has been received irrelevant to the goodness of the update in question. But clearly there are cases where it makes a lot of sense to privilege a later update over an earlier one: namely, where we have some substantive reason for doing so. I've emphasized throughout the difference between formal and substantive reasons for evidence. I've emphasized that my account is an account of the former. Nevertheless, it's important to have an idea of how my formal norm can be made consistent with the existence of substantive reasons. While I am again going to defer providing a much more worked out story than the one we already have on the table, I think we can at least say that the considerations that guide our instincts in the Jack and Jill case indicate that the second condition of Bayesian coherentism (Revised) is best understood as a *prima facie* constraint. Contra the way we have formulated Bayesian coherentism so far, then, the constraint on diachronic coherence encoded in EI should hold only in those cases where there are no substantive reasons for either of the updates that EI is defined over. This makes Bayesian coherentism consistent with the idea that, in some cases, the best possible credence distribution is incoherent by the lights of EI. This might be because we have substantive reason to favor one piece of evidence over another, as in the case where we have reason to favor the evidence given to us by Jill over the evidence given to us by Jack. Or, it might be because we have substantive reason to take both these updates seriously. In the latter case, being diachronically coherent will be a more complicated matter. The bottom line, however, is that it is only where there is no reason to do either of these things that EI kicks in to tell us something about the goodness of the belief revisions in question. Let us then revise Bayesian coherentism one last time:

> Bayesian coherentism (final): Where there is no substantive reason for either of a pair of probabilistic belief transitions, this pair of probabilistic belief transitions will be such that:

(a) There are sufficient partitions, $\{E_i\}$, $\{F_j\}$, for each of these updates, and

(b) The degree of diachronic incoherence of these updates is determined by EI.

Does revising our norm in this way make it objectionably weak? I think the answer to this question is that it makes Bayesian coherentism exactly as strong as we would want it to be, given the purpose for which it has been contrived. Recall our objective has been to get a norm capable of unifying Bayesian updates by replacing the foundationalist's constraint. But the foundationalist's constraint is itself remarkably weak. To really appreciate its weakness, consider an agent who, after being knocked over the head, directly changes her credence in some proposition to one, and then updates in accordance with her conditional probabilities. Such an update conforms to Bayesian foundationalism. Nevertheless, such an update clearly still goes wrong in an important way. The way in which it goes wrong has to do with the lack of substantive reason the agent has to revise her beliefs in the first place. The fact that Bayesian coherentism is also weak, in virtue of the updates it governs also lacking these reasons, is no objection to it then. Quite the opposite: it's exactly what we would expect of a formal norm for evidence.

I want to conclude by re-considering 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 ought to 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. For instance, maybe like Field (1978, p. 365) claims when discussing his own updating rule, we would want to say that, unlike Jeffrey conditioning, Bayesian conditioning is too much of an idealization to ever be of any use:

> I suspect that the fact that [Bayesian conditioning] is not a special case of [Field conditioning] is no loss—I suspect that [Bayesian conditioning] *should* be regarded as an oversimplification that can't ever really arise—but if you want to allow it, you can allow change to occur by [Bayesian conditioning] as well as by [Field conditioning].

Or maybe, like Lange (2000, p. 397), we would want to hold that the conditions under which updates on uncertain evidence happen differ in relevant ways from those under which updates on certain evidence happen:

> Whether the stimulus we receive succeeds in pushing our confidence in e to a given level in the open interval $(0,1)$ depends on our prior opinions. This does not arise in cases to which Bayesian conditioning applies, since then we would presumably have come away from our experience with $pr'=1$ whatever our prior level of confidence in e had been.

Lange does not elaborate on why he thinks only updates on uncertain evidence are sensitive to an agent's prior opinions. At one point, he notes 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", in the way that they do when we assign our evidence any other value.[24]

Despite the weird imagery, this does not seem like a crazy suggestion. For starters, it does seem as though some propositions, though they might be triggered by experience, are not *justified* by experience. When I change my credence in the proposition that a difficult math proof is correct, though this change may be accompanied by certain sensory experiences that are brought about by introspection, these experiences do not seem to be what *justify* these revisions, in the way that my belief that the sky is blue is justified by an experience with a certain phenomenal character.[25] If this is the case—and if it is also the case that updates on certain evidence are exactly those that aren't justified by experience—then the agent's prior expectations (her background beliefs) won't have a hand in determining what her experience justifies, since, in these cases, the agent's experience does 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 value of one might differ in some relevant way from the type of content to which we happen to be justified in assigning a lesser value. If there is indeed this difference in content between our certain and uncertain evidence, a unified account of the normative structure that this evidence partakes in 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 different types of propositions this evidence corresponds to. If so, then there may, after all, be reason to think that updates on certain and uncertain evidence ought to proceed from frameworks with different normative structures.[26]

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 cre-

[24] Lange (2000, p. 400).

[25] See Cassell (ms. a) for further discussion.

[26] Of course, this will depend upon how they are different. In the end, such a difference might very well turn out to be irrelevant as well.

dence of one. This might be a considerable task.[27] 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 the possible ways that the commitments underlying Bayesianism 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 epistemology. 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 the answer that's been provided besides its connection to the 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.

[27]It's true that we have principles that regulate content (think Lewis's (1980) Principal Principle and Van Fraassen's (1984) Rational Reflection). However, these principles are still formal, or 'syntactic', principles, in the sense that their prescriptions are based upon the formal relations that certain beliefs hold to certain other beliefs. Or, put another way, these principles hold for all beliefs, irregardless of their content. The Principal Principle maintains that you ought to calibrate your beliefs about certain propositions to your beliefs about the objective chances of those propositions, no matter what those propositions happen to be. Rational Reflection says that you ought to calibrate your beliefs about certain propositions to your beliefs about your future beliefs about those propositions, no matter what those propositions happen to be. In both cases, the constraint holds merely in virtue of the contents of two sorts of beliefs being identical. And identity is a formal relation.

By contrast, if we were looking for a principle to govern which beliefs ought to be assigned a value of one, this principle would have to target this content directly. It could not make use of a formal property, like identity.

Appendix A

There are different formal measures that might be used to give content to the notion of Bayesian evidential incoherence. In this appendix, I outline one such measure, which draws upon the seminal work of Diaconis and Zabell (1982) on the formal properties of Jeffrey conditioning.[28] This measure follows naturally from a property that is both necessary and sufficient for commutative updates (p.825):

> 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 $\{p_i\}$ 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.

It should be clear how the lack of Jeffrey independence undermines the commutativity of updates. Since Jeffrey updating fixes the probabilities on the evidence partition, the values for $\mathcal{F}$ that we get when we update on it first will be the same values for $\mathcal{F}$ that we get when we update on it second. If $P_{\mathcal{E}}(F_j)\neq P(F_j)$, then where we get $\mathcal{F}$ first and $\mathcal{E}$ second (where $P_{\mathcal{E}}=P_{\mathcal{F}\mathcal{E}}$), it will change the values along $\mathcal{F}$ from those that we got from updating on $\mathcal{F}$ first. Therefore, these values will differ from those that we would have gotten from updating on $\mathcal{F}$ second. Our final probabilities for the propositions in $\mathcal{F}$ will be different, then, depending upon whether we update on $\mathcal{F}$ first or second. Similarly, since Jeffrey updating fixes the probabilities on the evidence partition, the values for $\mathcal{E}$ that we get when we update on it first will be the same as the values for $\mathcal{E}$ that we get when we update on it second. If $P_{\mathcal{F}}(E_i)\neq P(E_i)$, then where we get $\mathcal{E}$ first and $\mathcal{F}$ second (where $P_{\mathcal{F}}=P_{\mathcal{E}\mathcal{F}}$), it will change the values along $\mathcal{E}$ from those that we got from updating on $\mathcal{E}$ first. Therefore, these values will differ from those that we would have gotten from updating on $\mathcal{E}$ second. Our final probabilities for the propositions in $\mathcal{E}$ will be different, then, depending upon whether we update on $\mathcal{E}$ first or second.

Since, as the authors show, Jeffrey independence is necessary and sufficient for commutativity (pp. 825-26), if we are looking to assess the degree to which some set of updates fails to commute, understanding what it would mean for this standard to fail to be met to varying degrees looks like the place to start. We can

[28] The following draws heavily from §3.3 of this discussion.

begin by distinguishing Jeffrey independence from a different, stronger type of independence (p. 825):

> P-INDEPENDENCE: Two partitions $\mathcal{E}=\{E_i\}$, and $\mathcal{F}=\{F_j\}$, such that $P(E_i)>0$, $P(F_i)>0$ for all *i* and *j*, are P-independent if $P(E_i \mid F_j)=P(E_i)$ and $P(F_j \mid E_i)=P(F_j)$ for all i, *j*.
>
> Thus, P-independence says that conditioning on $\mathcal{F}$ does not change the probabilities on $\mathcal{E}$ and similarly with $\mathcal{E}$ and $\mathcal{F}$ interchanged.

As should be clear, P-independence entails Jeffrey independence: it is Jeffrey independence, for all p_i, q_j. The reason that Jeffrey independence, rather than P-independence, is sufficient for commutativity is that, given two partitions whose members aren't all p-independent of each other, it is possible to update on these partitions and assign them values that perfectly offset these relations of dependence, in a way that secures the commutativity of updates. Therefore, the dependence that gives rise to non-commutativity—that which we are looking to capture, and to make gradable—will be a function, both of the relations of p-dependence that precede an update, and of the weighted evidence partition that gets updated on.

To begin to get a handle on this function, notice that since Jeffrey independence says that $P_{\mathcal{E}}(F_j) = P(F_j)$, for all *j*, and $P_{\mathcal{F}}(E_i) = P(E_i)$, for all *i*, it will hold where $\frac{P_{\mathcal{E}}(F_j)}{P(F_j)}=1$, for all *j*, and $\frac{P_{\mathcal{F}}(E_i)}{P(E_i)}=1$, for all *i*. Plausibly, then, the degree of a violation of Jeffrey independence will be a function of the amount by which each of these diverges from 1. To formulate a measure that can account for this in a perspicuous way, we can note, with Diaconis and Zabell, that $\frac{P_{\mathcal{E}}(F_j)}{P(F_j)}=\sum_i p_i r_{ij}$ where,

$$r_{ij} = \frac{P(E_iF_j)}{P(E_i)P(F_j)}.$$

Given this, the degree of a violation of Jeffrey independence will correspond to the sums of the amount by which $\sum_i p_i r_{ij}$ diverges from 1, for all *j*, and the amount by which $\sum_j q_j r_{ij}$ diverges from 1, for all *i*. These observations point us towards the following measure of incoherence:

> EVIDENTIAL INCOHERENCE (EI): Let P be a probability function and let $P_{\mathcal{E}}$ be the probability function that results from updating P on $\mathcal{E}$ with probabilities p_i. Further, let $r_{ij}=\frac{P(E_iF_j)}{P(E_i)P(F_j)}$, $r_{ij'}=\frac{P_{\mathcal{E}}(E_iF_j)}{P_{\mathcal{E}}(E_i)P_{\mathcal{E}}(F_j)}$, if defined, and 1 otherwise.

A sequence of updates over the partitions, $\{E_i\}$, $\{F_j\}$, is coherent to the extent that it minimizes $\sum_j(|1\text{-}\sum_i r_{*,j} p_i|)+\sum_i(|1\text{-}\sum_j r_{i,*'} q_j|)$.

We can run through a couple of examples to see how EI will work.

(1) Consider the following initial credence distribution $P(E_iF_j)$:

	F_1	F_2	F_3	
E_1	.25	.125	.125	.5
E_2	.125	0	.125	.25
E_3	.125	.125	0	.25
	.5	.25	.25	

Now assume that the agent gets as evidence $p(E_1)$=.5, $p(E_2)$=.2, $p(E_3)$=.3 and $p(F_1)$=.2, $p(F_2)$=.4, $p(F_3)$=.4, in turn. We are left with the following after we update on $\{E_i\}$, with values $\mathbf{p}_i$ (left), and on $\{F_j\}$ with values $\mathbf{q}_j$ (right):[29]

	F_1	F_2	F_3	
E_1	.25	.125	.125	**.5**
E_2	.1	0	.1	**.2**
E_3	.15	.15	0	**.3**
	.5	.275	.225	

	F_1	F_2	F_3	
E_1	.1	.182	.22	.502
E_2	.04	0	.18	.22
E_3	.06	.218	0	.278
	.2	**.4**	**.4**	

Before the first update, (r_{ij}) represents the relations of dependence that hold between $\{E_i\}$ and $\{F_j\}$. Before the second update, $(r_{ij'})$ represents the relations of dependence that hold between $\{E_i\}$ and $\{F_j\}$:

$$(r_{ij}) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix} \qquad (r_{ij'}) = \begin{pmatrix} 1 & .91 & 1.11 \\ 1 & 0 & 2.22 \\ 1 & 1.82 & 0 \end{pmatrix}$$

Multiplying each column (i.e., each member of $\{F_j\}$) of (r_{ij}) by $\mathbf{p}_i$, and taking the amount by which the value of each column diverges from one, and multiplying each row (i.e., each member of $\{E_i\}$) of $(r_{ij'})$ by $\mathbf{q}_j$, and taking the amount by which the value of each row diverges from one yields:

$$\sum_j(|1\text{-}\sum_i r_{*,j} p_i|)+\sum_i(|1\text{-}\sum_j r_{i,*'} q_j|)\approx.37$$

[29] Some of the values in the second chart are approximations. However, since they are not inputs into our measure, this makes no difference.

This is a measure of the evidential incoherence of the update.

(2) Now assume that, given the same initial credence distribution, the agent gets as evidence $p(\mathrm{E}_1)$=.6 $p(\mathrm{E}_2)$=.2, $p(\mathrm{E}_3)$=.2 and $p(\mathrm{F}_1)$=.4 $p(\mathrm{F}_2)$=.3, $p(\mathrm{F}_3)$=.3, in turn. We are left with the following after we update on $\{\mathrm{E_i}\}$, with values $\mathbf{p}_\mathrm{i}$ (left), and on $\{\mathrm{F_j}\}$ with values $\mathbf{q}_\mathrm{j}$ (right):

	F_1	F_2	F_3	
E_1	.3	.15	.15	**.6**
E_2	.1	0	.1	**.2**
E_3	.1	.1	0	**.2**
	.5	.25	.25	

	F_1	F_2	F_3	
E_1	.24	.18	.18	.6
E_2	.08	0	.12	.2
E_3	.08	.12	0	.2
	.4	**.3**	**.3**	

Before the first update, $(\mathrm{r_{ij}})$ represents the relations of dependence that hold between $\{\mathrm{E_i}\}$ and $\{\mathrm{F_j}\}$. Before the second update, $(\mathrm{r_{ij'}})$ represents the relations of dependence that hold between $\{\mathrm{E_i}\}$ and $\{\mathrm{F_j}\}$:

$$(\mathrm{r_{ij}})= \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix} \qquad (\mathrm{r_{ij'}})= \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$$

Multiplying each column (i.e., each member of $\{\mathrm{F_j}\}$) of $(\mathrm{r_{ij}})$ by $\mathbf{p}_\mathrm{i}$, and taking the amount by which the value of each column diverges from one, and multiplying each row (i.e., each member of $\{\mathrm{E_i}\}$) of $(\mathrm{r_{ij'}})$ by $\mathbf{q}_\mathrm{j}$, and taking the amount by which the value of each row diverges from one yields:

$$\textstyle\sum_\mathrm{j}(|1\text{-}\sum_\mathrm{i}\mathrm{r}_{*,\mathrm{j}}\mathrm{p_i}|)+\sum_\mathrm{i}(|1\text{-}\sum_\mathrm{j}\mathrm{r}_{\mathrm{i},*'}\mathrm{q_j}|)=0$$

This is a measure of the evidential incoherence of the update. This update exhibits perfect evidential coherence.

References

[1] BonJour, Laurence. 1988. *The Structure of Empirical Knowledge.* Cambridge: Harvard University Press.

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

[3] Bradley, Richard. 2005. "Radical Probabilism and Bayesian conditioning." *Philosophy of Science.*

[4] Carnap, Rudolf. 1950. "Empiricism, Semantics, and Ontology." *Revue Internationale de Philosophie* 4 (11):20-40.

[5] Cassell, Lisa. ms. a. "Jeffrey Conditioning: From Requantification to Dynamics".

[6] Cassell, Lisa. ms. b. "Commutativity, Normativity and Holism: Lange Revisited".

[7] Bradley, Richard. 2005. "Radical Probabilism and Bayesian conditioning." *Philosophy of Science.*

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

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

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

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

[12] Doring, Frank. 1999. "Why Bayesian Psychology is Incomplete". *Philosophy of Science* (Proceedings), 66: 379-S389.

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

[14] Feldman, Richard. 2001. "Voluntary Belief and Epistemic Evaluation." In Matthias Steup (ed.), *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*. Oxford University Press 77–92.

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

[16] Huemer, Michael, 2001. *Skepticism and the Veil of Perception* Lanham, MD: Rowman and Littlefield.

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

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

[19] 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.

[20] Keynes, John Maynard. 1921. *A Treatise on Probability.* Dover Publications.

[21] Kornblith, Hilary. 1992. "The laws of thought." *Philosophy and Phenomenological Research* 52 (4):895-911.

[22] Lange, Marc. 2000. "Is Jeffrey Conditioning Defective in Virtue of Being Non-Commutative." *Synthese*

[23] Laplace, P. S. "Essai Philosophique sure les Probabilites." Courcier, New York, 1814. English translation by FW Truscott and FL Emory.

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

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

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

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

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

[29] Weisberg, Jonathan. 2015. "Updating, Undermining and Independence." *British Journal for the Philosophy of Science*.

[30] Williams, Bernard. 1985. *Ethics and the Limits of Philosophy.* Harvard University Press.

[31] Wolterstorff, Nicholas. 2010. *Practices of Belief.* Cambridge University Press.