How to Learn from Theory-Dependent Evidence; or Commutativity and Holism: A Solution for Conditionalizers[*]

J. Dmitri Gallow

Final Draft. Forthcoming in *The British Journal for the Philosophy of Science*.

Abstract

Weisberg (2009) provides an argument that neither conditionalization nor Jeffrey conditionalization is capable of accommodating the holist's claim that beliefs acquired directly from experience can suffer undercutting defeat. I diagnose this failure as stemming from the fact that neither conditionalization nor Jeffrey conditionalization give any advice about how to rationally respond to *theory-dependent* evidence, and I propose a novel updating procedure which does tell us how to respond to evidence like this. This holistic updating rule yields conditionalization as a special case in which our evidence is entirely theory-independent.

1 Introduction

Weisberg's *Commutativity or Holism? A Dilemma for Conditionalizers* provides a compelling argument that neither of the orthodox belief-revision norms of partial belief epistemology—neither conditionalization nor Jeffrey conditionalization —is capable of accommodating the confirmational holist's claim that beliefs acquired directly from experience can suffer undercutting defeat. I will diagnose this failure as stemming from the fact that neither of these rules give any advice about how to rationally respond to experiences in which our evidence is theory-dependent, and I will propose a novel updating procedure which does tell us how to respond to these experiences. This holistic updating rule will be capable of properly modeling cases

[*]I am indebted to Yuval Abrams, James Joyce, Jason Konek, Maria Lasonen-Aarnio, Sarah Moss, Patrick Shirreff, Daniel J. Singer, Julia Staffel, Brian Weatherson, and anonymous reviewers for comments on earlier versions of this paper. Thanks also to audiences at the University of Michigan, the University of Southern California, and the City University of New York Graduate Center.

in which beliefs acquired directly from experience suffer undercutting defeat, and it will yield conditionalization as a special case in which our evidence is entirely theory-independent.

2 Conditionalization

Much of the discussion of holism has taken place against the backdrop of a *full belief* understanding of doxastic states, according to which an agent's doxastic state is given by a specification of the propositions which they believe, those which they disbelieve, and those which they neither believe nor disbelieve. However, there are many who think that doxastic states are richer than the full belief account gives them credit for. I believe both that global temperatures are rising and that I have hands, but the second belief is much stronger than the first. I believe that I have hands to a much greater *degree* than I believe that global temperatures are rising. The *partial belief* account attempts to capture these comparative features of my beliefs. It maintains that the fundamental doxastic state is a *credal state* which can be represented with a triple $< \mathcal{W}, \mathcal{P}, C >$ of a set of possible worlds $\mathcal{W}$, a set of propositions $\mathcal{P}$ (the propositions toward which the agent bears doxastic attitudes), and a credence function C which assigns numbers between 0 and 1 to the propositions in $\mathcal{P}$, where these numbers are interpreted as *degrees of belief* or *credences*.

We can lay down synchronic norms governing credal states at any time; a widespread position in partial belief epistemology is that credences should conform to the probability axioms—that C should be a probability function. Doxastic states, however, are not static. We frequently undergo learning experiences. So partial belief epistemology provides norms governing how a credal state ought to respond to such learning experiences. Suppose that we begin with a credence function C, and then undergo the learning experience $\mathcal{E}$. This should take us to a new credence function, which I'll denote '$C_{\mathcal{E}}$':

$$C \xrightarrow{\mathcal{E}} C_{\mathcal{E}}$$

What constraints should $C_{\mathcal{E}}$ satisfy? The orthodox Bayesian answer to this question depends in part upon the kind of learning experience that $\mathcal{E}$ is. For instance, suppose that the evidence gleaned from $\mathcal{E}$ (call that evidence '$\mathbb{E}$') is just a bunch of propositions, $\{e_1, ..., e_N\}$. In that case, orthodoxy says that $C_{\mathcal{E}}(\,\cdot\,)$ must be $C(\,\cdot \mid e_1...e_N)$, the prior credence function C *conditional* on the proposition $e_1...e_N$ (the conjunction of e_1 through e_N). This method of updating is known as *conditionalization*. Conditionalization says that if $\mathbb{E} = \{e_1, ..., e_N\}$, then[1]

$$C_{\mathcal{E}}(p) \stackrel{!}{=} C(p \mid e_1...e_N) \qquad (Condi)$$

Since $C(e_1...e_N \mid e_1...e_N) = 1$, conditionalizing upon the proposition $e_1...e_N$ is only appropriate if $\mathcal{E}$ rationalizes absolute certainty about the truth of $e_1...e_N$. Thus, $Condi$ presupposes that our evidence consists of just those propositions which experience has made it rational to believe with absolute certainty. Call pieces of evidence like this *propositional certainties*.

So long as the evidence acquired in experience consists entirely of propositional certainties, we ought to update our credences by conditionalization. However, we might think that there are learning experiences in which we gain evidence which is neither propositional nor certain. Richard Jeffrey (1965), for instance, believed that this could happen if, *e.g.*, we observe a color patch in low lighting for a short period. This experience might not tell us with any certainty whether the patch is green or blue, but it could nevertheless have the result of making us think that it is twice as likely to be green as blue—and such a reaction could be entirely rational. In general, Jeffrey believed that a learning experience could fail to tell us that a unique evidence proposition e is certain to be true, but could instead tell us that an n-tuple of evidence propositions $e_1, ..., e_n$ which partition $\mathcal{W}$ should be assigned an n-tuple of weights $\omega_1, ..., \omega_n$. We can represent this kind of evidence with a set of ordered pairs of propositions and weights $\mathbb{E} = \{< e_1, \omega_1 > ... < e_n, \omega_n >\}$ where the e_i partition $\mathcal{W}$ and the ω_i sum to 1. In virtue of this representation, call pieces of evidence like this *weighted partitions*. When experience provides a weighted partition, Jeffrey claimed that our posterior credence that p should be the weighted sum of our prior conditional credences that p given each of the e_i, with the weights given by the ω_i:

$$C_{\mathcal{E}}(p) \stackrel{!}{=} \sum_i C(p \mid e_i) \cdot \omega_i \qquad (JCondi)$$

This updating procedure is known as 'Jeffrey conditionalization.' Jeffrey's rule is not a *competitor* to $Condi$, but rather a *generalization* of $Condi$. $Condi$ only tells us what to do in very particular circumstances: circumstances in which the deliverances of

[1] A few words on notation: I'm placing a bang ! above the equals sign to indicate that the equality has normative, and not descriptive, force—the claim isn't that $C_{\mathcal{E}}(p)$ *will* be equal to $C(p \mid e_1...e_N)$, merely that it *should* be. Throughout 'p' will be used as a schematic variable ranging over propositions in $\mathcal{P}$, and I will represent the conjunction of p and q with 'pq'.

experience are propositional certainties. Evidence may come in different forms. We may receive evidence about the likelihood of each of several propositions—likelihoods which are not mediated by way of any proposition.[2] In that case, *Condi* is silent, so Jeffrey is free to step in and suggest a method for responding to such evidence without stepping on the toes of conditionalization. To further appreciate the compatibility of these two updating rules, note that *JCondi* has *Condi* as an instance. In the limiting case in which $\omega_j = 0$ for every $j \neq i$—in the limiting case in which we become certain that one of the propositions in $\{e_1, ..., e_n\}$ is true and all of the other propositions are false—Jeffrey conditionalization reduces to straight conditionalization and tells us to set our posterior credence in p to $C(p \mid e_i)$.

3 Holism and Conditionalization

Confirmational holism is the claim that the rational response to a learning experience can depend not just on the character of our experience, but additionally upon our prior doxastic state. For this reason, beliefs acquired directly from experience can suffer undercutting defeat. By way of explanation: suppose that you undergo an experience $\mathcal{E}$ which renders it rational to form the belief that e. Suppose further, in line with confirmational holism, that $\mathcal{E}$ only rationalized the belief that e because you accepted a certain background theory t. If you later get evidence that t is false, without having received any other reason for thinking that e in the interim, then you ought to lose your belief that e. If e was only allowed into the belief box because t vouched for it, then when t's credentials are called into question, so too are e's. In this case, your belief that e suffers undercutting defeat; it ceases to be rational to believe that e, not because you have been given new evidence which speaks against e, but rather because your prior grounds for accepting e have been undermined.

The straightforward way to square holism with *Condi* and *JCondi* is to say that which proposition or weighted partition a given learning experience provides us with depends in part upon our prior credence function C. On this approach, the rational response to a learning experience is always to update with either *Condi* or *JCondi*, but our background beliefs can affect which inputs we ought to feed into these update rules.

Jonathan Weisberg (2009), however, provides a compelling argument that this approach can't be made to work. More specifically, he shows that if we proceed this way, then we won't be able to capture cases in which beliefs acquired directly from

[2] Or perhaps: not mediated by way of any proposition included in $\mathcal{P}$

experience suffer undercutting defeat. To see why, consider the following case:

> **Trick Lighting**
> Sabeen has a visual experience as of a red cube in front of her ($\mathcal{E}_R$). Believing that the lighting in the room is normal (l_N), she forms the belief that the cube in front of her is red (c_R). She then looks up and has a visual experience as of a trick light ($\mathcal{E}_T$) which, she knows from her past experience with such lights, is capable of making objects appear any color whatsoever, no matter their actual color. She responds to this experience by losing her belief that the lighting in the room is normal and gaining the belief that it is trick lighting (l_T). She discards her belief that c_R and returns to a state of suspended judgment with respect to the color of the cube.

That's how the story goes on the full-belief account. Note that, according to the holist, the rational doxastic attitude for Sabeen to take with respect to the proposition c_R depends upon her beliefs about the lighting in the room. Upon experiencing $\mathcal{E}_R$, she ought to believe c_R only if she believes l_N. For this reason, when she later comes to lose the belief that l_N, she ought to lose the belief that c_R.

Holism is quite plausible in **Trick Lighting**. However, in the partial belief framework, it makes trouble for *Condi* and *JCondi*. In partial belief terms, holism requires that the degree to which Sabeen believes that c_R after having the experience $\mathcal{E}_R$ ought to depend upon the degree to which she believed that l_N before $\mathcal{E}_R$, and that, subsequent to $\mathcal{E}_R$, in the absence of any other reason to think that the cube is red, her credence that c_R ought to be lowered if her credence that l_N is lowered. The problem is that, unless Sabeen *started out* thinking that the kind of lighting in the room is relevant to the color of the cube (which, intuitively, she should not), neither *Condi* nor *JCondi* is capable of getting this result.

Before undergoing the experience $\mathcal{E}_R$, Sabeen ought to regard l_T and c_R as independent. She ought not think that the kind of lighting in the room provides any information about the color of the cube, nor that the color of the cube provides any information about the lighting. (To make it perfectly clear, we may suppose that Sabeen knows ahead of time that the kind of lighting in the room was determined by a die roll and the color of the cube was determined by a coin flip, and that the outcomes of those chance processes were causally independent of one another.) However, after undergoing the experience $\mathcal{E}_R$, Sabeen ought to regard l_T and c_R as dependent. Once she's had an experience as of a red cube, and raised her credence that the cube is red on the basis of that experience, any reason to think that that experience was illusory is a reason to think that her credence that c_R is too high and ought to be lowered—and l_T is a reason to think that her experience was illusory. On the orthodox story, it is only if l_T and c_R are dependent that the experience $\mathcal{E}_T$, which only directly carries information about the kind of lighting in the room, and not the color of the cube, could rationally lead Sabeen to lower her credence that c_R.

However, both *Condi* and *JCondi* are *rigid* updating procedures. That is, both *Condi* and *JCondi* preserve conditional credences $C(p \mid e_i)$ for each e_i which is either conditioned upon (for the case of *Condi*) or which belongs to the weighted partition (for the case of *JCondi*). Thus, for any proposition p, updating by *Condi* or *JCondi* cannot introduce dependence between p and any of the e_i which was not already present in Sabeen's prior credence function.[3] Since Sabeen should start off regarding c_R and l_T as independent, and since $\mathcal{E}_R$ only carries information about the color of the cube (Sabeen's experience as of a red cube doesn't provide any evidence about the kind of lighting in the room, even if what evidence it does provide *depends* upon her believing that the lighting is normal), the experience $\mathcal{E}_R$ is incapable of introducing the necessary dependence between c_R and l_T—so long as we take Sabeen's evidence to be limited to c_R or a weighting of the $\{c_R, \overline{c}_R\}$[4] partition, and so long as Sabeen updates her beliefs using either *Condi* or *JCondi*. And if c_R and l_T remain independent, then Sabeen's credence that c_R will not change when she undergoes the experience $\mathcal{E}_T$. For that experience only directly affects her credence that l_T, and changing her credence that l_T could only have an effect on her credence that c_R if $C_{\mathcal{E}_R}(c_R \mid l_T) \neq C_{\mathcal{E}_R}(c_R)$, which won't be the case if Sabeen responded to $\mathcal{E}_R$ by updating her beliefs about the color of the cube using *Condi*, with the proposition that c_R, or *JCondi*, with some weighting of the partition $\{c_R, \overline{c}_R\}$.

We might think that Sabeen shouldn't be (Jeffrey) conditionalizing on the proposition c_R, but rather on a proposition like *the cube appears red*. Very well; the conditionalizer is free to pick their epistemological ground floor. However, no matter which ground floor they pick, the confirmation holist will think that there are cases in which *that* ground floor is susceptible to undercutting defeat. For illustration, have Sabeen condition on the proposition that the cube appears red—we can denote that proposi-

[3] Note that this is not to say that *Condi* and *JCondi* cannot introduce dependence between *any* propositions which were independent in the prior credence function. They both can. It is merely to say that they cannot do so for the epistemic ground-floors—the propositions or the cells of the weighted partitions—which are fed into those rules.

[4] Throughout, I'll use '$\overline{p}$' for the negation of p.

tion 'c_R^α'. Now, after conditionalizing on c_R^α, Sabeen might gain evidence that she is a test subject in a rogue neuroscientist's experiment, and that they have slipped her a drug which makes her an unreliable judge of color appearances. Call this proposition 'd' (for 'drug'). Before having the experience $\mathcal{E}_R$, Sabeen ought to regard c_R^α and d as independent. Prior to the experience, information about whether she's been slipped the drug doesn't tell her anything about how the cube appears, nor does information about how the cube appears tell her anything about whether or not she's been slipped the drug. However, afterward, she ought to regard them as dependent, for exactly the reasons discussed above. But neither *Condi*, with the proposition c_R^α, nor *JCondi*, with a weighting of the $\left\{c_R^\alpha, \overline{c_R^\alpha}\right\}$ partition, is capable of getting this result, for exactly the reasons discussed above. The move to appearance propositions just moves the bump in the carpet. (I'll have more to say on the move to appearance propositions at the end of section 4 below.)

You might think that Sabeen really *shouldn't* end up seeing the evidence that she's been slipped the drug as telling her anything about whether or not the cube appears red. You might think that there is an epistemological ground floor which can never suffer undercutting defeat. Be that as it may. Weisberg's goal is not to persuade you that confirmational holism is correct—merely that it is inconsistent with *Condi* and *JCondi*. Similarly, my goal is simply to persuade you that, while *Condi* and *JCondi* are not consistent with holism, there is a way of extending *Condi* to cover cases involving theory-dependent evidence which *is* consistent with holism. You need not accept confirmational holism in order to accept either of these theses.

4 A Holistic Update

Thus far, I've been running with the idea that the right way to make sense of theory-dependent evidence in the partial belief framework is to say that experiences give us the kind of evidence presupposed by *Condi* and *JCondi*—either a propositional certainty or a weighted partition—but that *which* such piece of evidence it gives us depends upon our prior credal state. Here's another way to make sense of theory-evidence dependence: we suppose that learning experiences can provide us with, not a propositional certainty, nor a weighted partition of propositions, but rather a dependence relation between background theories and evidence propositions. On this way of making sense of theory-evidence dependence, Sabeen's experience doesn't just flat-out give her the evidence that the cube is red. Rather, it gives her evidence with caveats—it tells her that if the lighting is normal, then her evidence is that the cube is red. And it tells her that if the lighting is trick, then she doesn't have any evidence at all about the color of the cube.

If we represent this kind of dependence between a background theory t and a piece of evidence e with an ordered pair $< t, e >$, then the present suggestion is that, sometimes at least, the rational import of an experience $\mathcal{E}$ can be represented as a relation between background theory and evidence, $\{< t_1, e_1 > ... < t_N, e_N >\}$. (Since the background theories t are themselves propositions, this is just a binary relation between propositions in $\mathcal{P}$.) What it is for an experience to provide a theory-evidence dependence relation like this is for it to be such that, for each i, if t_i is true, then the experience provides the evidence that e_i. For instance, in the **Trick Lighting** case, Sabeen's experience $\mathcal{E}_R$ provides her with the following relation between theory and evidence:

$$\{< l_N, c_R >, < l_T, \top >\}$$

If the lighting is normal, then Sabeen receives the evidence that the cube is red. If, however, the lighting is trick, then Sabeen merely receives the trivial evidence $\top$.[5] That is to say, if the lighting is trick, then she does not receive any evidence at all.

In the back of my mind as I write is a Williamsonian theory of evidence, according to which evidence is just knowledge and, since knowledge is not luminous (we are not always in a position to know what we know), evidence is not luminous either (we are not always in a position to know what our evidence is).[6] Therefore, what our evidence is can vary depending upon which background theory of ours is true. It then makes sense to think of raw experience as underdetermining our evidence; there are states of the world compatible with our experience in which we know that e and there are states of the world compatible with our experience in which we fail to know that e. We are not always in a position to know which of these states of the world actually obtains. We are therefore not always in a position to know what our evidence is. However, while experience doesn't determine evidence, it does determine a mapping from possible states of the world to potential evidence. A Williamsonian could understand my theory-evidence dependence relations as providing this kind of mapping. If t_i is true, then the experience $\mathcal{E}$ puts you in a position to know that e_i.[7]

[5] '$\top$' denotes the set of all possible worlds, $\mathcal{W}$.

[6] *C.f.* Williamson (2000, ch. 9)

[7] We may not always be in a position to know that our raw experience provides us with one dependence relation, as opposed to some other dependence relation. No matter; we needn't claim that we will always be in a position to know how we ought to proportion our beliefs in order to make claims about how we ought to proportion them. Indeed, if Williamson is correct about anti-luminosity, then *no* epistemology

By the way, on other accounts of evidence, it would be wrong to call the propositions which are dependent upon the background theories 'evidence'. However, I don't think that disputes over the nature of evidence should make much difference to my suggestion for how the confirmational holist ought to formally represent the deliverances of experience, nor to my suggestion (given below) for how to rationally respond to these deliverances.[8] One is free to call the experience $\mathcal{E}$, the relation $\{< t_1, e_1 > \ldots < t_N, e_N >\}$, or the propositions e_i 'evidence'. It won't make any difference to the formal representation or the rational response.

What *is* the rational response to an experience like this? We might think that we should just conditionalize on the material conditional $t_i \supset e_i$, for each i.[9] However, this leads to bad results in **Trick Lighting**. To make the case concrete, suppose that Sabeen starts out with credence 0.9 that the lighting in the room is normal and credence 0.1 that the lighting in the room is trick, and that she initially divides her credence equally between the cube being red and the cube being green (c_G). Suppose further that Sabeen regards the lighting in the room to be independent of the color of the cube, and suppose that she has no more reason to think that the cube will appear red if the lighting is trick than she has reason to think that the cube will appear green if the lighting is trick.[10] Then, Sabeen's prior doxastic state will be the one displayed on the left-hand side of

---

will be capable of providing a norm such that we will always be in a position to know whether we are following that norm (*c.f.* Hawthorne and Srinivasan 2013). In those cases in which we fail to proportion our beliefs correctly because we're not in a position to know which dependence relation our experience provides, we will fall short of ideal rationality, though perhaps blamelessly.

On a related note: the reader might worry that such an understanding of theory-evidence dependence is inherently anti-holistic, since *which* theory-evidence dependence relation an experience provides is not up for revision. Note, firstly, that I did not define holism à la Quine, as the view that everything in the web of belief is up for revision. I defined it as the view that, sometimes at least, the beliefs formed directly on the basis of experience can suffer undercutting defeat. Nothing in this thesis should commit us to the claim that every belief is rationally revisable. Moreover, even if we *did* accept this stronger Quinean thesis, it wouldn't entail that the theory-evidence dependence relation should be revisable, since that relation is independent of what Sabeen does or ought to believe about it. She needn't have any beliefs about that relation at all. If she does, then perhaps those beliefs could be rationally revised. But that wouldn't mean that the theory-evidence dependence relation itself had been revised, anymore than Sabeen's revised beliefs about the cube's color mean that the cube itself has changed color.

[8] Substantive views about the nature of evidence will bear on the question of which theory-evidence dependence relation a given experience provides. For the purposes of this paper, I remain neutral on such questions. My goal is not to formulate necessary and sufficient conditions for an experience providing a particular theory-evidence dependence relation, but rather to provide a formal framework for representing and rationally responding to theory-evidence dependence.

[9] This approach is implicitly suggested by Wagner's treatment of Weisberg's argument (*c.f.* Wagner 2013).

[10] Also, let's stipulate that Sabeen is certain that the trick light will not make the cube appear any color besides red or green.

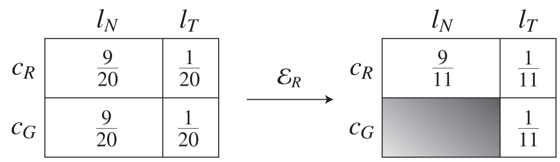


Figure 1: How Sabeen's credences change if she conditions on $l_N \supset c_R$

figure 1. (By the way, if Sabeen is at all reasonable, she'll save some of her credence for the proposition that the cube is a color other than red or green, as well as setting some credence aside for the proposition that she's been slipped the drug which undermines the reliability of her appearance judgments, that she's a brain-in-a-vat, and so on and so forth. In choosing to model Sabeen's rational doxastic state as I have, I am simply supposing that the amount of credence Sabeen leaves for these propositions is so paltry as to be negligible, and am therefore neglecting it.)

Given this specification of Sabeen's credal state, the result of conditionalizing on the material conditional $l_N \supset c_R$ is shown on the right-hand side of figure 1. Notice that, while conditionalizing on this conditional will raise her credence that the cube is red and render the propositions c_R and l_T dependent, as it ought, it will also raise her credence that the lighting is trick, as it ought not. If she antecedently thought that the trick lighting was more likely to make the cube appear red than green, or that the cube was more likely to be green than red, then it would make sense to raise her credence that the lighting is trick in response to an experience as of a red cube. However, Sabeen thought that the cube was just as likely to be red as green, and she had no more reason to think that the trick lighting would make the cube appear red than she had reason to think that it would make the cube appear green. So an experience as of a red cube shouldn't give her any reason to suspect that the lighting isn't normal.[11]

So conditionalizing on the material conditional $t \supset e$ doesn't work. Though, on reflection, we shouldn't find this too surprising; the situation we are considering is one in which Sabeen's experience tells her that, if the lighting is normal, then her evidence is that the cube is red. But that is not equivalent to her experience giving her the evidence that if the lighting is normal then the cube is red. In the former case, the evidence is relativized to a background theory. In the latter case, the evidence is just another run-of-the-mill proposition. It is important here to clearly distinguish two claims: the first

[11] As the reader may verify for themselves, Sabeen's credence that l_T will also be raised if she uses Jeffrey conditionalization to shift her credence over the partition $\{l_N \supset c_R, l_N c_G\}$.

is that your evidence is that if t is true, then e.

$$\mathbb{E} = \{t \supset e\}$$

The second is that, if t is true, then your evidence is e.

$$t \ \supset \ \mathbb{E} = \{e\}$$

I am claiming that the latter, and not the former, is the right way to understand theory-evidence dependence.

If conditionalizing on the material conditional doesn't work, what does? Here is a suggestion: when an agent receives the theory-evidence dependence relation $\{< t_1, e_2 > \dots < t_N, e_N >\}$, they should partition their credal state by the background theories t_i, and, *within* each t_i, proceed as if they had the evidence e_i. Since, if they have the evidence e_i, the thing to do is to conditionalize upon it, they ought to conditionalize upon the proposition e_i within the background theory t_i. Here is a general rule which will achieve this:

$$C_{\mathcal{E}}(p) \stackrel{!}{=} \sum_i C(p \mid t_i \, e_i) \cdot C(t_i) \qquad (HCondi)$$

If an experience provides an agent with a theory-evidence dependence relation $\{< t_1, e_1 > \dots < t_N, e_N >\}$ such that the t_i partition $\mathcal{W}$, then they should update in the manner specified by *HCondi*.[12] This is equivalent to the claim that their posterior credence function should be a probability which satisfies the following constraints,

[12] In the body, I presuppose that the background theories upon which the evidence depends will partition $\mathcal{W}$—that is, that they will both *cover* $\mathcal{W}$ (every possible world will be a member of at least one of the t_i) and that they will be *disjoint* (no world will be a member of more than one of the t_i). We shouldn't in general expect that either of these properties will be satisfied by the background theories which determine the evidential import of experiences. Fortunately, even if the background theories don't partition $\mathcal{W}$, we can transform any theory-evidence dependence relation $\{< t_1, e_1 > \dots < t_N, e_N >\}$ into an equivalent theory-evidence dependence relation $\{< t'_1, e'_1 > \dots < t'_M, e'_M >\}$ such that the t'_i *do* partition $\mathcal{W}$. To see how to do that, first consider the case in which our experience tells us that, if t_1, then our evidence is e_1, and if t_2, then our evidence is e_2, but t_1 and t_2 are not disjoint. Then, we can replace $\{< t_1, e_1 >, < t_2, e_2 >\}$ with $\{< t_1 \bar{t}_2, e_1 >, < t_1 t_2, e_1 e_2 >, < \bar{t}_1 t_2, e_2 >\}$. In this way, we end up with an equivalent theory-evidence dependence relation such that the theories are disjoint. This can be extended to the case of $n > 2$ overlapping theories in a straightforward manner. Second, consider the case in which the background theories t_i fail to cover $\mathcal{W}$. Then, take the uncovered portion of $\mathcal{W}$, $\bigwedge_i \overline{t_i}$, and stipulate that, in that case, your evidence is just the necessary proposition $\top$. That is, include the additional ordered pair $< \bigwedge_i \overline{t_i}, \top >$ in the relation. Using these two methods, we can take any theory-evidence dependence relation whatsoever and transform it into an equivalent theory-evidence dependence relation such that the theories partition $\mathcal{W}$.

for every i:

$$C_{\mathcal{E}}(p \mid t_i) = C(p \mid t_i\, e_i)$$
$$C_{\mathcal{E}}(t_i) = C(t_i)$$

Note that, in the special case in which there's no theory-evidence dependence, and our experience just delivers the evidence e, *sans* caveat, we can let the background theory be the necessary proposition $\top$ and represent the deliverance of experience with $\{< \top, e >\}$. In this special case, *HCondi* reduces to *Condi*.

$$\begin{aligned} C_{\mathcal{E}}(p) &= C(p \mid \top\, e) \cdot C(\top) \\ &= C(p \mid e) \end{aligned}$$

Therefore, *HCondi* is consistent with *Condi* in exactly the way that *JCondi* is consistent with *Condi*: it has *Condi* has a special case. This allows us to embrace the above account without forsaking *Condi*, or having to explain away the myriad arguments for that belief-revision rule. We just have to understand that that rule, and those arguments, only apply in cases in which the deliverance of experience consists entirely of propositional certainties. They do not apply in cases in which the evidential import of a learning experience is theory-dependent, just as they do not apply to the kinds of learning experiences which Jeffrey was concerned with.

To illustrate how *HCondi* handles **Trick Lighting**: suppose that Sabeen divides her credence equally between c_R and c_G and has credence 0.9 that the lighting is normal and credence 0.1 that the lighting is trick, just as above. Suppose also that she has no more reason to think that the trick light will make the cube appear red than she has reason to think that it will make the cube appear green, just as above. On the current proposal, we can represent the deliverance of her experience $\mathcal{E}_R$ with $\{< l_N, c_R >, < l_T, \top >\}$. Assuming the lighting is normal, her evidence is that the cube is red. If, however, the lighting is trick, then $\mathcal{E}_R$ doesn't give Sabeen any evidence about the color of the cube. If she conforms to *HCondi*, then she'll respond to $\mathcal{E}_R$ by updating as follows:

$$\begin{aligned} C_{\mathcal{E}_R}(c_R) &= C(c_R \mid l_N\, c_R) \cdot C(l_N) \;\; + \;\; C(c_R \mid l_T\, \top) \cdot C(l_T) \\ &= 0.9 + 0.5 \cdot 0.1 \\ &= 0.95 \end{aligned}$$

Her credence that the cube is red shoots up from 0.5 to 0.95. Moreover, updating with

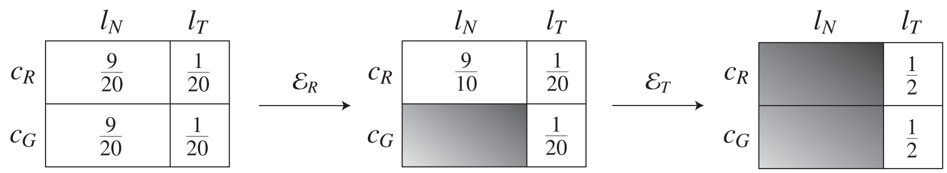


Figure 2: How Sabeen ought to update her credences according to *HCondi*

HCondi renders c_R and l_T dependent,

$$C_{\mathcal{E}_R}(c_R \mid l_T) = 0.5 \neq C_{\mathcal{E}_R}(c_R),$$

as figure 2 makes clear. However, unlike conditionalizing on the material conditional $l_T \supset c_R$, updating with *HCondi* will not have the bad result of raising Sabeen's credence that the lighting is trick. $C_{\mathcal{E}_R}(l_T)$ will be $C(l_T)$, just as it ought.

Now, suppose that Sabeen learns that l_T. For simplicity's sake, suppose that this evidence is theory-independent, so that Sabeen ought to update using *Condi*. Conditionalizing upon l_T will now have the effect of bringing Sabeen's credence that the cube is red back down to 0.5:

$$C_{\mathcal{E}_R,\mathcal{E}_T}(c_R) = C_{\mathcal{E}_R}(c_R \mid l_T) = 0.5$$

Her original reason for believing that c_R (to such a high degree) has been undercut by the defeater l_T, just as it ought.

As I mentioned above, the defender of orthodoxy might want to interject at this point by suggesting that, even though *HCondi* gives us the correct posterior distribution, *Condi* was capable of getting this posterior distribution on its own. After all, above, I had to specify that Sabeen had no more reason to think that the trick lighting would make the cube appear red than she had reason to think it would make the cube appear green. It was that stipulation that made trouble for conditionalizing on the material conditional $l_N \supset c_R$, since with that stipulation in place, it would be irrational for Sabeen to respond to $\mathcal{E}_R$ by getting more confident that the lighting is trick. However, with that stipulation in place, *Condi* is capable of arriving at precisely the same posterior distribution as *HCondi* by just updating on the proposition that the cube appears red, c_R^α, as shown in figure 3.

True enough.[13] But, to reiterate: to say this is merely to move the bump in the

[13] Bracketing some worries about the implicit appeal to the principle of indifference ('POI'). (Sabeen

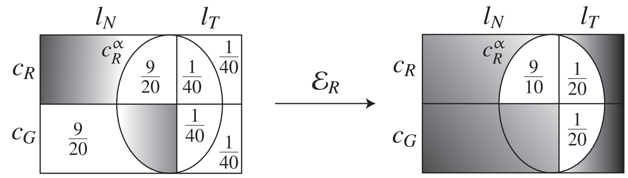
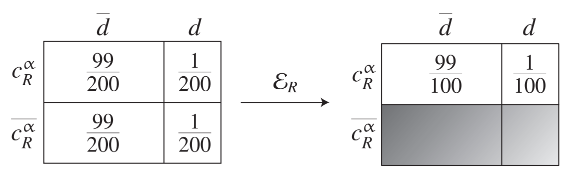


Figure 3: The proposition c_R^α is true inside the oval and false outside

carpet. The defender of conditionalization is free to pick their epistemological ground floor. Whatever ground floor they pick, the holist will claim that there are cases in which *that* ground floor will be susceptible to undercutting defeat. Above, I made the choice to neglect the negligible credence that Sabeen ought to allocate to the proposition that the rogue neuroscientist has slipped her a drug which renders her appearance judgments unreliable (d). Let me stop neglecting that credence. Suppose that, prior to undergoing $\mathcal{E}_R$, she's 99% confident that she hasn't been slipped the drug; she has no more reason to think that drug would make the cube appear to appear red than she has to think that it will make the cube appear to appear green; and she takes the appearance of the cube to be independent of whether or not she's been slipped the drug. Then, conditionalizing on the proposition c_R^α will do the following to Sabeen's doxastic state:



Thus, even though conditionalizing on the appearance proposition c_R^α will put Sabeen in a position to respond appropriately to the undercutting defeater l_T, it will not put her in a position to respond appropriately to the undercutting defeater d. With this posterior credal state, Sabeen still takes c_R^α to be independent of d. She still fails to see d as a reason to revise her views about whether or not the cube appears red. This problem remains whether she straight conditionalizes on c_R^α or merely readjusts her credences on the $\left\{ c_R^\alpha, \overline{c_R^\alpha} \right\}$ partition with *JCondi*. Nor does conditionalizing on the material conditional $\overline{d} \supset c_R^\alpha$ help, since that has the bad result of making Sabeen more confident that she has been slipped the drug:

may not have any more reason to think that the trick light will make the cube appear red than she has reason to think that it will make the cube appear green. However, this doesn't imply that she should divide her credence equally between these propositions, unless we assume the POI.)

$$\begin{array}{c|c|c|} & \overline{d} & d \\ \hline c_R^\alpha & \frac{99}{200} & \frac{1}{200} \\ \hline \overline{c_R^\alpha} & \frac{99}{200} & \frac{1}{200} \\ \hline \end{array} \xrightarrow{\mathcal{E}_R} \begin{array}{c|c|c|} & \overline{d} & d \\ \hline c_R^\alpha & \frac{99}{101} & \frac{1}{101} \\ \hline \overline{c_R^\alpha} & & \frac{1}{101} \\ \hline \end{array}$$

But, so long as Sabeen didn't antecedently have any more reason to think that the drug would make the cube appear to appear red than she had reason to think that it would make the cube appear to appear green, $\mathcal{E}_R$ shouldn't give her any reason to think that she's been slipped the drug.

Now, you might think that Sabeen's rational credence in the appearance proposition c_R^α really *shouldn't* end up dependent upon her credence that she's been slipped the drug. To reiterate: my goal here isn't to persuade you that there are cases in which the beliefs acquired directly from experience suffer undercutting defeat. My goal is just to persuade you that *Condi* and *JCondi* are inconsistent with this claim, and that *HCondi* is not.

It is instructive to see how *HCondi* handles the same problem. Once we stop neglecting the credence Sabeen gives to the proposition d, there are three background theories upon which Sabeen's evidence depends: the theory that the lighting is normal and she hasn't been given the drug, $l_N\overline{d}$, the theory that the lighting is trick and she hasn't been given the drug, $l_T\overline{d}$, and the theory that she's been given the drug, d. If $l_N\overline{d}$, then $\mathcal{E}_R$ gives the evidence that c_R, as well as the evidence that c_R^α. If $l_T\overline{d}$, then it merely gives the evidence that c_R^α. If d, then $\mathcal{E}_R$ doesn't give any evidence at all.[14] Then, the deliverance of $\mathcal{E}_R$ is representable as $\{< l_N\overline{d}, c_R c_R^\alpha >, < l_T\overline{d}, c_R^\alpha >, < d, \top >\}$, and *HCondi* will tell Sabeen to revise her beliefs as shown in figure 4. (In figure 4, by the way, I'm supposing that, initially, Sabeen is 99% confident that she wasn't slipped the drug; that she is 90% confident that the lighting is normal; that she divides her credence equally between both c_R and c_G and c_R^α and $\overline{c_R^\alpha}$; and that she takes each of these sets of propositions to be independent of the others.)

This posterior distribution cannot be achieved by conditionalizing on c_R, nor c_R^α, nor $(l_N\overline{d} \supset c_R) \wedge (l_T\overline{d} \supset c_R^\alpha)$, which will have the bad result of making Sabeen more confident that the lighting is trick and that she's been slipped the drug. The defender of *Condi* might attempt the short-sighted strategy of appealing to the proposition *the cube appears to appear red* ($c_R^{\alpha\alpha}$). If $c_R^{\alpha\alpha}$ is what we learn directly from experience, then the holist will take there to be a potential undercutting defeater for it, and neither

[14] I'm supposing that d is also a defeater for c_R.

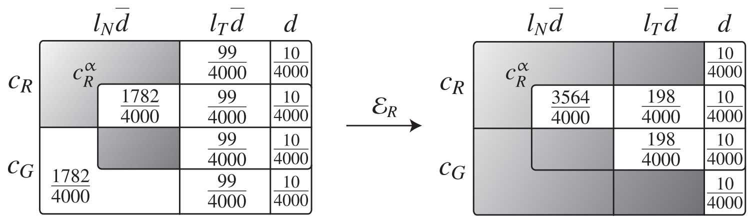


Figure 4: How Sabeen ought to update her beliefs according to *HCondi*, with the potential undercutting defeater d included (c_R^α is true inside the interior rectangle, and false outside).

Condi nor *JCondi* will adequately prepare Sabeen for this defeater, for the very same reasons discussed above. This game could go on forever, but at no point would the conditionalizer start winning.

5 HCondi and Dutch Books

The reader may be wondering whether responding to a theory-evidence dependence relation like $\{< l_N, c_R >, < l_T, \top >\}$ by updating with *HCondi* will leave Sabeen vulnerable to a so-called 'Dutch book', a series of wagers, the combination of which is guaranteed to lose money no matter what. The answer is: no, it will not. Responding by conditionalizing on the material conditional $l_N \supset c_R$, on the other hand, *will* leave her susceptible to a Dutch book.

After undergoing $\mathcal{E}_R$, the strongest proposition of which Sabeen becomes certain is $l_N \supset c_R$. Nevertheless, her posterior credal function $C_{\mathcal{E}_R}$ fails to satisfy the constraint that

$$C_{\mathcal{E}_R}(l_T) = C(l_T \mid l_N \supset c_R)$$

And, it might be thought, the Lewis-Teller Diachronic Dutch Book Argument for conditionalization demonstrates that if Sabeen violates this condition, then she could be sold a series of wagers which are guaranteed to lose money no matter what (Teller 1973). For instance, given the initial credal state shown in figure 2, Sabeen will see each of the following wagers as fair:

Wager 1

\$1800	if l_T
−\$400	if $l_N c_R$
\$0	if $l_N c_G$

Wager 2

\$81	if $l_N \supset c_R$
−\$99	if $l_N c_G$

(The numbers on the left represent Sabeen's net gain from the wager if the proposition to the right turns out to be true.) And, if Sabeen updates on $\{< l_N, c_R >, < l_T, \top >\}$ using *HCondi*, then she'll end up seeing the following wager as fair:

Wager 3

−\$1980	if l_T
\$220	if l_N

The Lewis-Teller argument proceeds as follows: a clever bookie could look at these dispositions and sell Sabeen a series of wagers which will lose her money come what may. First, he would have her agree to wagers 1 and 2 before learning anything. Then, if she updates on the theory-dependent evidence $\{< l_N, c_R >, < l_T, \top >\}$ using *HCondi*, she'll become certain that $l_N \supset c_R$ is true and therefore certain that $l_N c_G$ is false. At that point, have her agree to Wager 3. If, on the other hand, she becomes certain that $l_N c_G$ is true, there's no need to sell her Wager 3; you can just take her \$99 straight away. In this manner, you would be guaranteed to take \$99 from Sabeen, *no matter what*, as the following table demonstrates:

	l_T	$l_N c_R$	$l_N c_G$
Wager 1	\$1800	−\$400	\$0
Wager 2	\$81	\$81	−\$99
Wager 3	−\$1980	\$220	×
Net Profit	−\$99	−\$99	−\$99

(The table shows *Sabeen's* profit.)

Applied to theory-dependent evidence, the clever bookie's strategy is not so clever. It presupposes that Sabeen will *either* become certain that $l_N \supset c_R$ *or* she'll become certain that $l_N c_G$. This is what allows us to put a ×, rather than a \$220, in the final column next to Wager 3—the bookie's strategy was to not sell the third wager in the event that Sabeen becomes certain that $l_N c_G$. But in **Trick Lighting**, Sabeen will not, under any contingency, become certain that $l_N c_G$. We know from the get-go that Sabeen will either acquire the theory-evidence dependence relation $\{< l_N, c_R >, < l_T, \top >\}$ or the dependence relation $\{< l_N, c_G >, < l_T, \top >\}$.[15] In neither of these cases will updating with *HCondi* make Sabeen certain that $l_N c_G$.

[15] Or at least, if we filled in all the details in the appropriate way, we could make it so that we know this from the get-go.

Concocting a diachronic Dutch book strategy requires knowing ahead of time which possible learning experiences the agent might undergo. In the Lewis-Teller Dutch book strategy, it is assumed that we know ahead of time that the agent will learn *whether* p (for some proposition p). That is, they will either become certain that p or they will become certain that $\overline{p}$. This assumption is not valid in cases like **Trick Lighting**, where Sabeen will either become certain that $l_N \supset c_R$ or become certain that $l_N \supset c_G$.

So Sabeen can't be hoodwinked with the standard Lewis-Teller diachronic Dutch book. On the other hand, the straight conditionalizer who responds to $\{< l_N, c_R >, < l_T, \top >\}$ by conditionalizing on the material conditional $l_N \supset c_R$ *will* be susceptible to a diachronic Dutch book. Since experience will either provide the theory-evidence dependence relation $\{< l_N, c_R >, < l_T, \top >\}$ or $\{< l_N, c_G >, < l_T, \top >\}$, this agent will either conditionalize on $l_N \supset c_R$ or they will conditionalize on $l_N \supset c_G$. In either case, their credence that l_N will drop from $9/10$ to $9/11$. Before they have the experience, you will be able to sell them a \$110 bet on l_N for \$99. Afterwards, no matter which experience they undergo, you'll be able to buy it back from them for \$90. You'll walk away \$9 richer, no matter what.

6 Commutativity and Learning about Background Theories

I've argued that *HCondi* handles cases like **Trick Lighting** better than either $Condi$ or $JCondi$. However, **Trick Lighting** was a rather sterile case in certain respects. For instance, I had to stipulate that Sabeen didn't have any more reason to think that the trick lighting would make the cube appear red than she had reason to think that it would make the cube appear green. Without this supposition, it wouldn't have been untoward for her to become more confident that the lighting was trick after having an experience as of a red cube. Nevertheless, we can consider other, less sterile cases. Suppose that Sabeen thought that the trick lighting was twice as likely to make the cube red as to make it appear not red. That is, suppose that she was in the doxastic state shown on the left-hand side of figure 5. Then, presumably, she *ought* to see the experience $\mathcal{E}_R$ as providing a reason to think that the lighting is trick. However, if she updates with *HCondi* on the theory-evidence dependence relation $\{< l_N, c_R c_R^\alpha >, < l_T, c_R^\alpha >\}$, her credence that l_T will stay fixed. She'll end up at the posterior credal state shown on the right-hand side of figure 5.

Additionally, applications of *HCondi* will not in general commute. That is, with *HCondi*, updating first on the theory-evidence dependence relation E_1 and then on

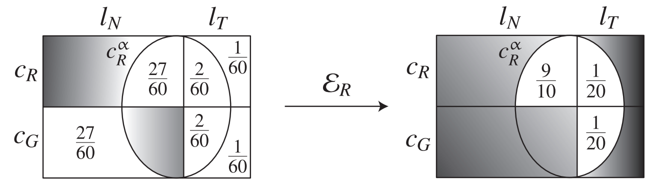


Figure 5: *HCondi* doesn't allow background theories to be confirmed by evidence

E_2 is not guaranteed to take you to the same credal state that you would have reached by updating first on E_2 and then on E_1. To see this, consider a credal state involving six propositions, t_1, t_2, t_3, e_1, e_2, and e_3. On almost any distribution over these propositions, *HCondi* applied to $E_1 = \{< t_1 \vee t_2, \overline{e_1} >, < t_3, \top >\}$ and $E_2 = \{< t_1, \top >, < t_2 \vee t_3, \overline{e_3} >\}$ will not commute (one such distribution is shown in figure 6).

Condi and *JCondi* weren't capable of getting the simple case right, so *HCondi*'s failing in the more complicated cases shouldn't prompt us to go running back to them. What we have to do is find a way of extending *HCondi* so that it is commutative and so that it allows background theories to be confirmed by the evidence.

6.1 Commutativity

Let me deal with commutativity first. To set the stage: there are two ways of thinking about the belief revision norm *Condi*. On one picture, you walk around with your current credence function C, and when you receive a new set of evidence $\mathbb{E}$, you conditionalize on that evidence to get a posterior credence function C', and then you might as well throw the prior credence function and the evidence away. Everything they have to tell you has already been encoded in your updated credence function C'. On the other picture, at any moment t, you have a body of total cumulative evidence $\mathbb{E}_t$ and an *initial* credence function C. The credal state you ought to have at t is then just your initial credence function conditional on your total cumulative evidence, $C(\cdot \mid \mathbb{E}_t)$. For *Condi*, so long as you never lose any evidence, it doesn't make any difference which of these pictures you adopt. However, for *HCondi*, it does. If we adopt the first picture, then *HCondi* will not be commutative. If we adopt the second, however, *HCondi* can be made to commute.

Here's how: suppose that we have two theory-evidence dependence relations $E = \{< t_1, e_1 > ... < t_N, e_N >\}$ and $E' = \{< t'_1, e'_1 > ... < t'_M, e'_M >\}$. Then, we can define a conjunction operation, $E \wedge E'$, which takes these two relations to a new relation

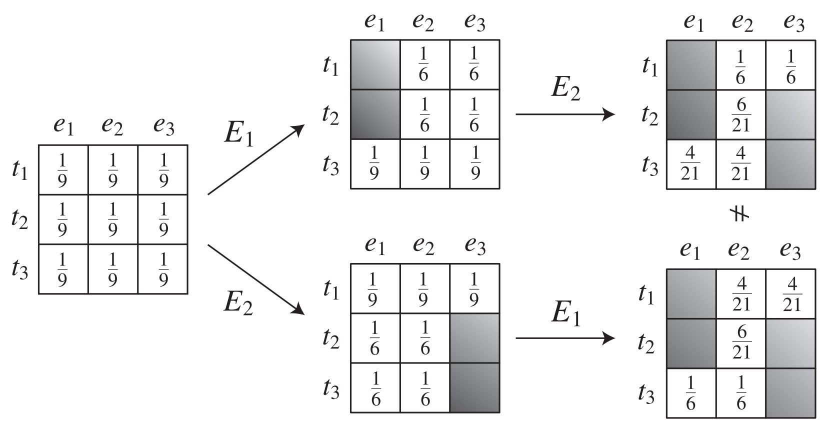


Figure 6: *HCondi* does not commute. With *HCondi*, the result of updating on the theory-evidence dependence relations $E_1 = \{< t_1 \vee t_2, \overline{e_1} >, < t_3, \top >\}$ and $E_2 = \{< t_1, \top >, < t_2 \vee t_3, \overline{e_3} >\}$ depends upon the order in which we update.

which encodes the theory-evidence dependences of both, like so:

$$E \wedge E' = \left\{ \begin{array}{cccc} < t_1 t'_1, e_1 e'_1 > & < t_1 t'_2, e_1 e'_2 > & \cdots & < t_1 t'_M, e_1 e'_M > \\ < t_2 t'_1, e_2 e'_1 > & < t_2 t'_2, e_2 e'_2 > & \cdots & < t_2 t'_M, e_2 e'_M > \\ \vdots & \vdots & \ddots & \vdots \\ < t_N t'_1, e_N e'_1 > & < t_N t'_2, e_N e'_2 > & \cdots & < t_N t'_M, e_N e'_M > \end{array} \right\}$$

When we conjoin two theory-evidence dependence relations in this way, if, for any theories t_i and t_j, $t_i t_j = \perp$, then we can safely throw out the ordered pair $< t_i t_j, e_i e_j >$. It only tells us what our evidence would be in an impossible scenario; since we are certain such a scenario will never arise, $< t_i t_j, e_i e_j >$ has nothing interesting to tell us. If $e_i e_j = \perp$, then we should set $C(t_i t_j)$ to zero and renormalize. If $e_i e_j = \perp$, then we've learned that if the background theory $t_i t_j$ is true, then our evidence is contradictory; and this provides a *reductio* of that background theory.

Now, we can say that at any moment t, the agent will have an initial credence function C and a set of theory-evidence dependences $E_t = \bigwedge_i E_i$, the conjunction of all the theory-dependent evidence that the agent has collected at time t, and the rational credal state to have at t is just the result of updating C with E_t, in the manner specified by *HCondi*.[16] Since the conjunction operator $\wedge$ is commutative, this method

[16] In the final analysis: the rational credal state at t is the result of updating C with E_t, in the manner

of updating will commute.

6.2 Learning about Background Theories

This solution to commutativity only deepens our first problem: that *HCondi* is an inherently segregative rule; it places walls between background theories and prevents credence from crossing those walls. This leads to the problem that, when two background theories t_i and t_j agree that the agent has acquired the evidence $e_i \vee e_j$, and t_i did a better job predicting this evidence than t_j, t_i fails to get any credit for its good prediction. Since my solution to *HCondi*'s noncommutativity was, in effect, to put up even more walls, it has rendered this first problem all the more pressing.

In this section, I'm going to suggest a way to extend *HCondi* so that we can adjust the credence of the background theories in light of the evidence. Recall that *HCondi* is equivalent to the constraint that your posterior credence function $C_{\mathcal{E}}$ both be a probability and that, for every i,

$$C_{\mathcal{E}}(p \mid t_i) = C(p \mid t_i\, e_i)$$
$$C_{\mathcal{E}}(t_i) = C(t_i)$$

I'm going to suggest that, in order to allow agents to learn about background theories, we replace the second of these constraints with

$$C_{\mathcal{E}}(t_i) = C(t_i) \cdot \Delta_i$$

where Δ_i tells us how the agent's credence that t_i ought to change in response to the experience $\mathcal{E}$. Δ_i will be a non-negative number; if it is greater than 1, then t_i is confirmed by the experience, and $C_{\mathcal{E}}(t_i)$ should be greater than $C(t_i)$. If Δ_i it less than 1, then t_i is disconfirmed by the experience, and $C_{\mathcal{E}}(t_i)$ should be less than $C(t_i)$. If $\Delta_i = 1$, then the experience neither confirms nor disconfirms t_i, and $C_{\mathcal{E}}(t_i)$ should equal $C(t_i)$.

Note that both *Condi* and *JCondi* have this general form. For *Condi*, we set $C_{\mathcal{E}}(t_i)$ equal to $C(t_i) \cdot \Delta_i$, where

$$\Delta_i = \frac{C(e \mid t_i)}{C(e \mid \top)}$$

That is: when we conditionalize upon an evidence proposition e, there is a baseline

specified by *HCondi*[*], defined below.

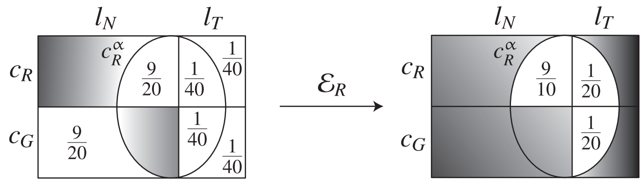


Figure 7: The result of updating on $\{< l_N, c_R c_R^\alpha >, < l_T, c_R^\alpha >\}$ with $HCondi$

likelihood for e, $C(e \mid \top)$. If a theory makes e *more* likely than this baseline, then it gets credit for its good prediction, and its credence goes up ($\Delta_i > 1$). If a theory makes e *less* likely than the baseline, then it gets punished for its poor prediction, and its credence goes down ($\Delta_i < 1$). If the theory makes e just as likely as the baseline, then it is neither rewarded nor punished, and its credence stays constant ($\Delta_i = 1$).

Similarly, for $JCondi$, $C_{\mathcal{E}}(t_i) = C(t_i) \cdot \Delta_i$, where

$$\Delta_i = \sum_j \frac{C(e_j \mid t_i)}{C(e_j \mid \top)} \cdot \omega_j$$

And I'm going to suggest that we do essentially the same thing in the case of $HCondi$. However, there is a wrinkle that must be dealt with. In the kinds of learning experiences covered by $Condi$ and $JCondi$, all of the background theories agree about what information the agent received from their experience. However, in the cases covered by $HCondi$, the background theories can disagree about what evidence the experience has provided. For instance, in the model of **Trick Lighting** displayed in figure 7, if l_N is correct, then Sabeen has acquired the evidence $c_R c_R^\alpha$. And l_N made this piece of evidence significantly more likely than did l_T.

$$C(c_R c_R^\alpha \mid l_N) = 1/2 > 1/4 = C(c_R c_R^\alpha \mid l_T)$$

However, l_T does not concur. According to it, she only received the evidence that c_R^α. And both l_N and l_T made this evidence equally likely.

$$C(c_R^\alpha \mid l_N) = C(c_R^\alpha \mid l_T) = 1/2$$

Different background theories end up disagreeing about how well they did predicting the evidence, in virtue of disagreeing about what the evidence is. The question of how to update the background theories t_i in the light of the evidence is therefore a question

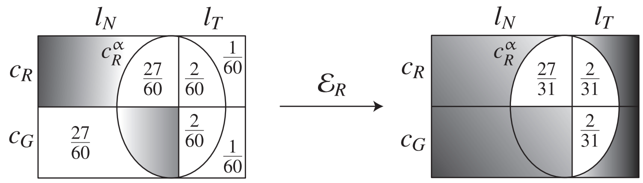


Figure 8

of diplomacy. How should Sabeen broker a truce between these disputing theories?

She ought not allow each background theory to conditionalize upon its own purported piece of evidence. That will result in l_N taking credence away from l_T in the case shown in figure 7. But, just as Sabeen ought not get more confident that the lighting is *trick* after undergoing $\mathcal{E}_R$, she also ought not get more confident that the lighting is *normal*.

I suggest that Sabeen proceed the way every good diplomat proceeds: seek out common ground. l_N and l_T may disagree about whether or not c_R is part of her evidence, but they agree that c_R^α is part of her evidence. Let them exchange credence on the basis of this common-ground evidence proposition. In the case shown in figure 7, this won't make any difference (they both did equally well predicting c_R^α). However, there are other cases in which it will make a difference. Consider, for instance, the credal state shown on the left-hand side of figure 8. There, Sabeen starts out believing that the trick light is twice as likely to make the cube appear red as not. Given this, she ought to end up seeing the experience as of a red cube, $\mathcal{E}_R$, as giving her some information about the lighting in the room. If, in this case, she sets $C_\mathcal{E}(l_T)$ and $C_\mathcal{E}(l_N)$ to $C(l_T \mid c_R^\alpha)$ and $C(l_N \mid c_R^\alpha)$, then she will end up more confident that the lighting is trick after undergoing $\mathcal{E}_R$ (as shown on the right-hand side of figure 8).

Just because l_N and l_T agree that Sabeen has learned that c_R^α, this doesn't mean that *every* background theory will agree that Sabeen has learned that c_R^α. For instance, in a model including the defeater proposition d (that Sabeen has been slipped a drug which impairs her color appearance judgments), d will not agree that c_R^α is part of Sabeen's evidence. However, this shouldn't keep l_N from spotting l_T some credence for what it recognizes as a good prediction.

Here's a way to let l_T and l_N work out their business without getting d involved: first consider the set of *agreed upon* evidence propositions, $\varepsilon = \{\varepsilon_1 ... \varepsilon_T\}$. If two or more background theories agree that Sabeen has learned that ε_i, then ε_i is included in ε. And consider the set of corresponding unions of agreeable background theories

$\tau = \{\tau_1 ... \tau_T\}$. For every ε_j, there is a unique τ_j such that, if a background theory t_i agrees that the agent has learned that ε_j, then $t_i \subset \tau_j$. (Note that the τ_i's needn't be disjoint, since a background theory t_i could agree with t_j that $e_i \vee e_j$ was learned, and also agree with t_k that $e_i \vee e_k \neq e_i \vee e_j$ was learned. Then, there will be two unions of agreeable background theories, τ_m and τ_n, such that $t_i \subset \tau_m$ and $t_i \subset \tau_n$.) The τ_i provide us with baseline likelihoods for the agreed-upon evidence, $C(\varepsilon_1 \mid \tau_1) ... C(\varepsilon_T \mid \tau_T)$, relative to which particular background theories can do better or worse. If they do better, then they should be rewarded with an increase in credence, proportional to the antecedent probability of τ_i; if they do worse, they should be punished with a decrease proportional to the antecedent probability of τ_i. We can accomplish this if we let Δ_i be

$$\Delta_i \equiv_{df} \sum_j \frac{\delta(\varepsilon_j \mid t_i)}{C(\varepsilon_j \mid \tau_j)} \cdot \frac{C(\tau_j)}{\sum_k C(\tau_k)}$$

where

$$\delta(\varepsilon_j \mid t_i) \equiv_{df} \begin{cases} C(\varepsilon_j \mid \tau_j) & \text{if } t_i \not\subset \tau_j \\ C(\varepsilon_j \mid t_i) & \text{if } t_i \subset \tau_j \end{cases}$$

The reason for the δ function is that we don't want a theory t_i to get an undue boost in credence simply because it agrees with a lot of other theories about what the evidence was, and therefore has more summands in its Δ_i term than other background theories. To avoid this, we include all the same summands in every t_i's Δ_i. This gives us a single basis along which to measure the (dis)confirmation of every background theory. If a theory doesn't think that ε_j was learned, then it is simply given the baseline score $C(\varepsilon_j \mid \tau_j)$ for that summand.

Here's an analogy for what's going on here that will hopefully make things clearer: We bring together all the background theories in a large conference hall. There are several negotiation tables labeled τ_1, τ_2, ..., τ_T. On each table τ_i is a proposition ε_i, and a baseline credence $C(\varepsilon_i \mid \tau_i)$—the credence given to ε_i by the union $\tau_i = \bigcup_j t_j$ of every t_j such that t_j agrees that ε_i was learned. Every background theory sits down at table τ_i, and they are each given a fraction $C(\tau_i)/\sum_j C(\tau_j)$ of their total prior credence. If the background theory doesn't think that ε_i was learned, then it keeps its $C(\tau_i)/\sum_j C(\tau_j)^{th}$ of credence and sits tight. If, however, the theory agrees that ε_i was learned, then, if it made ε_i *less* likely than the baseline, it must hand some of its credence back, in proportion to how unlikely it made ε_i. If a theory agrees that ε_i was learned and it made ε_i *more* likely than the baseline, then it gets to take some credence from those who did worse, in proportion to how likely it made ε_j. They go from table to table like this, until they have sat at every table. Whatever credence they have at the end of this process is theirs to keep. So, if a theory t_i doesn't agree with any of the other theories about what was learned, it will walk away with its same initial credence $C(t_i)$ ($\Delta_i = 1$). If a theory agrees with some other theories about what evidence was learned, and it made all of these evidence propositions more likely than their baselines, then its credence will go up ($\Delta_i > 1$). If a theory agrees with some other theories about what evidence was learned, and it made all of these evidence propositions less likely than their baselines, then its credence will go down ($\Delta_i < 1$).

We can now put forward the fully general version of *HCondi*—let's call it '*HCondi**':

$$C_{\mathcal{E}}(p) \stackrel{!}{=} \sum_i C(p \mid t_i\, e_i) \cdot C(t_i) \cdot \Delta_i \tag{$HCondi^*$}$$

It is easy to verify that *HCondi** reduces to *HCondi* in the cases we considered in section 4. By the way, it's far from obvious that, with the Δ_is included, $C_{\mathcal{E}}$ will still be a probability function. Fortunately, it turns out that it will be; in the appendix, I supply a proof.

7 In Summation

Following Weisberg (2009), I have argued that, according to holism, properly modeling cases like **Trick Lighting**—cases in which beliefs acquired directly from experience suffer undercutting defeat—requires that an updating procedure can render such evidence dependent on background beliefs which the agent previously took to be independent of that evidence. If that is so, then neither *Condi* nor *JCondi* is capable of delivering the correct verdict about these cases, since both *Condi* and *JCondi* are rigid updating procedures. They preserve conditional credences $C(p \mid e_i)$ for all propositions p and all evidence propositions e_i. My diagnosis of this situation was that neither *Condi* nor *JCondi* give advice about how to respond to theory-dependent evidence. And I proposed a novel updating rule, *HCondi*, which does tell us what to do with evidence like this. This updating rule gets cases like **Trick Lighting** right, and agrees with *Condi* in the special case in which our evidence is theory-independent. I went on to explain why *HCondi* will not render an agent diachronically Dutch-bookable; I provided an understanding of *HCondi* on which it commutes; and I extended the rule so that it could allow agents to learn about the background theories which determine the evidential import of their experiences.

Appendix

$C_{\mathcal{E}}$ is a Probability. *If $C(p)$ is a probability function, then*

$$C_{\mathcal{E}}(p) = \sum_i C(p \mid t_i, e_i) \cdot C(t_i) \cdot \Delta_i$$

where

$$\Delta_i \equiv_{df} \sum_j \frac{\delta(\varepsilon_j \mid t_i)}{C(\varepsilon_j \mid \tau_j)} \cdot \frac{C(\tau_j)}{\sum_k C(\tau_k)}$$

and

$$\delta(\varepsilon_j \mid t_i) \equiv_{df} \begin{cases} C(\varepsilon_j \mid \tau_j) & \text{if } t_i \not\subset \tau_j \\ C(\varepsilon_j \mid t_i) & \text{if } t_i \subset \tau_j \end{cases}$$

will be a probability function as well.

Proof. The non-negativity and finite additivity of $C_{\mathcal{E}}$ follows from the nonnegativity and finite additivity of C. The proofs are trivial.

To prove normality, I will need the following lemma:

Lemma 1.

$$\sum_i C(t_i) \cdot \Delta_i = 1$$

Proof. Pulling out the common factor, each Δ_i consists of T summands:

$$\Delta_i = \frac{1}{\sum_k C(\tau_k)} \cdot \left(\frac{\delta(\varepsilon_1 \mid t_i)}{C(\varepsilon_1 \mid \tau_1)} \cdot C(\tau_1) + \frac{\delta(\varepsilon_2 \mid t_i)}{C(\varepsilon_2 \mid \tau_2)} \cdot C(\tau_2) + \ldots + \frac{\delta(\varepsilon_T \mid t_i)}{C(\varepsilon_T \mid \tau_T)} \cdot C(\tau_T) \right)$$

Denote the j^{th} summand of Δ_i with 'Δ_i^j'. Then,

$$\Delta_i = \frac{1}{\sum_k C(\tau_k)} \left(\Delta_i^1 + \Delta_i^2 + \ldots + \Delta_i^T \right)$$

and

$$\begin{aligned} C(t_i) \cdot \Delta_i &= \frac{1}{\sum_k C(\tau_k)} \left(C(t_i) \cdot \Delta_i^1 + C(t_i) \cdot \Delta_i^2 + \ldots + C(t_i) \cdot \Delta_i^T \right) \\ &= \frac{1}{\sum_k C(\tau_k)} \cdot \sum_j C(t_i) \cdot \Delta_i^j \end{aligned}$$

We therefore get that

$$\begin{aligned}\sum_i C(t_i)\cdot\Delta_i &= \frac{1}{\sum_k C(\tau_k)}\sum_i\sum_j C(t_i)\cdot\Delta_i^j\\ &= \frac{1}{\sum_k C(\tau_k)}\sum_j\sum_i C(t_i)\cdot\Delta_i^j\end{aligned}$$

For an arbitrary τ_j, and an arbitrary $t_i \subset \tau_j$,

$$\begin{aligned}C(t_i)\cdot\Delta_i^j &= \frac{C(\varepsilon_j \mid t_i)}{C(\varepsilon_j \mid \tau_j)}\cdot C(t_i)\cdot C(\tau_j)\\ &= \frac{C(\varepsilon_j \mid t_i)}{C(\varepsilon_j \mid \tau_j)}\cdot\frac{C(t_i)}{C(\tau_j)}\cdot C(\tau_j)^2\\ &= \frac{C(\varepsilon_j, t_i, \tau_j)}{C(t_i, \tau_j)}\cdot\frac{C(t_i, \tau_j)}{C(\tau_j)}\cdot\frac{C(\tau_j)}{C(\varepsilon_j, \tau_j)}\cdot C(\tau_j)^2\\ &= C(t_i \mid \tau_j, \varepsilon_j)\cdot C(\tau_j)^2\end{aligned}$$

And summing over all $t_i \subset \tau_j$,

$$\sum_i C(t_i)\cdot\Delta_i^j = \sum_i C(t_i \mid \tau_j, \varepsilon_j)\cdot C(\tau_j)^2 = C(\tau_j)^2\cdot\sum_i C(t_i \mid \tau_j, \varepsilon_j) = C(\tau_j)^2$$

Now, for an arbitrary $t_i \not\subset \tau_j$,

$$C(t_i)\cdot\Delta_i^j = \frac{C(\varepsilon_j \mid \tau_j)}{C(\varepsilon_j \mid \tau_j)}\cdot C(t_i)\cdot C(\tau_j) = C(t_i)\cdot C(\tau_j)$$

So, summing over all $t_i \not\subset \tau_j$,

$$\sum_i C(t_i)\cdot\Delta_i^j = \sum_i C(t_i)\cdot C(\tau_j) = C(\tau_j)\cdot\sum_i C(t_i) = C(\tau_j)\cdot C(\overline{\tau}_j)$$

Putting together the $t_i \subset \tau_j$ and the $t_i \not\subset \tau_j$, therefore, we get that

$$\sum_i C(t_i)\cdot\Delta_i^j = C(\tau_j)^2 + C(\tau_j)\cdot C(\overline{\tau}_j) = C(\tau_j)\cdot\big(C(\tau_j) + C(\overline{\tau}_j)\big) = C(\tau_j)$$

Since j was arbitrary,

$$\frac{1}{\sum_k C(\tau_k)} \sum_j \sum_i C(t_i) \cdot \Delta_i^j = \frac{1}{\sum_k C(\tau_k)} \sum_j C(\tau_j) = 1$$

Thus,

$$\sum_i C(t_i) \cdot \Delta_i = 1$$

establishing the lemma $\square$

With **Lemma 1** in hand, normality follows immediately:

$$\begin{aligned} C_{\mathcal{E}}(\top) &= \sum_i C(\top \mid t_i, e_i) \cdot C(t_i) \cdot \Delta_i \\ &= \sum_i C(t_i) \cdot \Delta_i \\ &= 1 \end{aligned}$$

$\square$

References

Christensen, D. (1992). Confirmational Holism and Bayesian Epistemology. *Philosophy of Science 59*(4), 540–557.

Hawthorne, J. and A. Srinivasan (2013). Disagreement without Transparency: Some Bleak Thoughts. In J. Lackey and D. Christensen (Eds.), *The Epistemology of Disagreement: New Essays*, pp. 9–30. Oxford: Oxford University Press.

Jeffrey, R. (1965). *The Logic of Decision*. New York: McGraw-Hill.

Teller, P. (1973). Conditionalization, observation, and change of preference. In W. L. Harper and C. A. Hooker (Eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Volume I, pp. 205–253. Dordrecht: D. Reidel Publishing Company.

Wagner, C. (2013). Is Conditioning Really Incompatible with Holism? *Journal of Philosophical Logic 42*(2), 409–414.

Weisberg, J. (2009). Commutativity or Holism? A Dilemma for Conditionalizers. *British Journal for the Philosophy of Science 60*(4), 793–812.

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