On the Proper Formulation of Conditionalization

Michael Rescorla

Abstract: Conditionalization is a norm that governs the rational reallocation of credence. I distinguish between factive and non-factive formulations of Conditionalization. Factive formulations assume that the conditioning proposition is true. Non-factive formulations allow that the conditioning proposition may be false. I argue that non-factive formulations provide a better foundation for philosophical and scientific applications of Bayesian decision theory. I furthermore argue that previous formulations of Conditionalization, factive and non-factive alike, have almost universally ignored, downplayed, or mishandled a crucial *causal* aspect of Conditionalization. To formulate Conditionalization adequately, one must explicitly address the causal structure of the transition from old credences to new credences. I offer a formulation of Conditionalization that takes these considerations into account, and I compare my preferred formulation with some prominent formulations found in the literature.

§1. Conditionalizing on a falsehood

The central notion of Bayesian decision theory is *credence* (a.k.a. *degree of belief*, a.k.a. *subjective probability*). Bayesians study an idealized agent who attaches credences to propositions. The agent's credence in proposition H is given by $P(H)$, a real number between 0 and 1. According to Bayesians, credences should conform to axioms of the probability calculus. The axioms set norms for rational credal allocation. Also crucial to the Bayesian framework is *conditional probability* $P(H \mid E)$, the subjective probability of H given E, which is standardly defined as follows:

(1) $$P(H \mid E) =_{df} \frac{P(H \,\&\, E)}{P(E)}.$$

I confine attention to cases where $P(E) > 0$, so that (1) is well-defined.[1]

Suppose you begin with an initial credal allocation P_{old} and subsequently adopt a new credal allocation P_{new}. You have *conditionalized on proposition E* when

(2) $P_{new}(\,.\,) = P_{old}(\,.\mid E)$.

A *conditioning proposition* is a proposition on which you conditionalize. Using (1), we can convert (2) into a more informative expression for $P_{new}(H)$:

(3) $$P_{new}(H) = \frac{P_{old}(H \,\&\, E)}{P_{old}(E)}.$$

Bayesians endorse a norm called *Conditionalization*, which requires you to conditionalize on E in certain circumstances. Ramsey (1931/1990, p. 88) offers an early statement:

> Since an observation changes… my opinion about the fact observed, some of my degrees of belief after the observation are necessarily inconsistent with those I had before. We have therefore to explain how exactly the observation should modify my degrees of belief; obviously if *p* is the fact observed, my degree of belief in *q* after the observation should be equal to my degree of belief in *q* given *p* before, or by the multiplication law to the quotient of my degree of belief in *pq* by my degree of belief in *p*.

Earman (1992, p. 34) gives a representative modern formulation:

> The rule of *strict conditionalization* says that if it is learned for sure that E and if E is the strongest such proposition, then the probability functions Pr_{old} and Pr_{new}, representing

---

[1] I discuss cases where $P(E) = 0$ in (Rescorla, forthcoming).

> respectively degree of belief prior to and after acquisition of new knowledge, are related by $\mathrm{Pr}_{\mathrm{new}}(.) = \mathrm{Pr}_{\mathrm{old}}(./E)$.

And here's Joyce (2009, p. 419):

> Imagine a person whose state of uncertainty is characterized by a "prior" probability P_0 …, and who is not dogmatic about x, so that $1 > P_0(x) > 0$. If the person undergoes a learning experience in which the only new information she acquires is that x is certainly true, then her post-learning "posterior" probability P_1 should coincide with her pre-learning probability conditional on x, $P_1(\bullet) = P_0(\bullet|x)$.

Ramsey, Earman, and Joyce use different language to describe the relation between agent and conditioning proposition. For Ramsey, the agent "observes" a "fact." For Earman, she "learns a proposition for sure," thereby acquiring new "knowledge." For Joyce, she "acquires the information that x is certainly true." All three formulations imply, or strongly suggest, that the conditioning proposition is true. In that sense, all three formulations are *factive*.

Many authors adopt or at least intimate a factive approach to Conditionalization, including Arntzenius (2003, p. 367, fn. 7), Briggs (2009, p. 61), de Finetti (1970/1974, p. 141), Eagle (2011, p. 117), Easwaran (2011, p. 314), Greaves and Wallace (2006, p. 607), Hacking (1967, p. 314), Howson and Franklin (1994, p. 453), Jeffreys (1948, pp. 29-40), Lewis (1999, p. 405), Savage (1954/1972, pp. 43-50), Skyrms (1987, pp. 2-3), Vineberg (2011), Weisberg (2009, p. 499), and Williamson (2000, pp. 184-223). In contrast, van Fraassen (1999, p. 93) offers a resolutely *non-factive* formulation:

> Our opinion consists of a probability function; new evidence consists of propositions to which we give probability 1… [T]here is a Modus Ponens like operation, *Conditionalization*, which counts as purely logical updating. If P is my prior opinion, and

E my new evidence, then P′ = P(|E), defined as P(&E)/P(E), is my new (posterior) opinion --- the change of P into P′ is called *conditionalization on* E.

Since an agent may assign credence 1 to a false proposition, van Fraassen's formulation does not require that E be true. Hájek (2009, p. 165), Huber (2016), Jeffrey (2004, p. 51), Talbott (2015), Titelbaum (2013, p. 52) also formulate Conditionalization non-factively.

Which is preferable --- a factive or non-factive formulation? The literature does not adequately address this question. Few authors explicitly distinguish factive from non-factive formulations, let alone provide reasons favoring one approach over the other. Conditionalization is central to philosophical and scientific applications of Bayesian decision theory, so we should resolve this matter in order to formulate Conditionalization as optimally as possible.

I contend that non-factive formulations offer a decisive advantage over factive formulations: they cover a wider range of important cases. Factive formulations do not cover scenarios involving *misplaced certainty*: i.e. scenarios where an agent sets $P_{new}(E) = 1$ yet E is false. Misplaced certainty is possible, so we would like to isolate rational norms that govern it. We therefore require a suitable non-factive formulation of Conditionalization. That is what the present paper aims to provide.

§§2-3 deploy the phenomenon of misplaced certainty to motivate non-factive Conditionalization. §4 offers my preferred non-factive formulation. §§5-6 elucidate my formulation and compare it with some prominent alternatives. I argue that previous formulations, factive and non-factive alike, have almost universally ignored, downplayed, or mishandled a crucial *causal* aspect of Conditionalization. To formulate Conditionalization adequately, one must explicitly address the causal structure of the transition from P_{old} to P_{new}. §7 illustrates my approach by examining scenarios that involve *memory loss*. §8 highlights key features of Conditionalization that emerge from my discussion.

§2. Misplaced certainty

For any proposition E with $P(E) > 0$, it is an easy theorem of the probability calculus that $P(E \mid E) = 1$. If $P_{new}(\,.\,) = P_{old}(\,.\mid E)$, and if P_{new} conforms to the probability calculus axioms, then $P_{new}(E) = 1$. Thus, conditionalizing on E requires you to assign credence 1 to E. Bayesians commonly use the phrase

X is certain that p,

stipulatively to mean

X assigns credence 1 to p.

Under this usage, conditionalizing on E requires you to be certain of E.

Factive formulations of Conditionalization restrict attention to scenarios where $P_{new}(E) = 1$ and E is true. Given virtually any such scenario, we can envision an altered scenario where $P_{new}(E) = 1$ and E is false. Humans are fallible. All ordinary channels of epistemic justification --- including perception, testimony, memory, introspection, and *a priori* reasoning --- are subject to error. No matter how compelling your grounds for setting $P_{new}(E) = 1$, those grounds are virtually always consistent with E's falsity. For example, suppose that perceptual experience leads you to set P_{new}(*I have hands*) = 1. As the skeptical tradition makes abundantly clear, you still may not have hands.

Perhaps we should restrict attention to conditioning propositions E that concern your own present mental states (e.g. *It looks to me as if I have hands*)? These propositions are more plausible candidates for infallible knowledge, so one might argue that within this domain $P_{new}(E) = 1$ guarantees *E*'s truth.

Arguably, though, you are fallible even about your own present mental states (Williamson, 2000, pp. 93-113, pp. 164-183). More importantly, propositions about your own present mental states do not form an adequate basis for philosophical and scientific applications of the Bayesian framework. Most applications feature conditioning propositions that concern matters far beyond one's immediate mental states. For example, Bayesian philosophy of science models the scientist as updating credences based upon observational or experimental data. The "data" is invariably expressed through propositions that concern the external world: readings from ammeters, scales, timers, and other measuring instruments; economic variables (e.g. interest rates or price levels); geographic patterns of fossil distribution; morbidity rates in various populations; and so on. Scientists, along with Bayesian philosophers of science, treat external world propositions as the primary evidential basis for scientific theorizing. In contrast, propositions about one's own mental states play no serious evidentiary role within modern science (excepting certain areas of psychology that rely on introspection). Any credible Bayesian philosophy of science must allow conditioning propositions that concern the external world.

A similar diagnosis applies to most other philosophical and scientific applications of the Bayesian framework. Two examples:

- *Bayesian statistics* uses Bayesian norms to guide statistical inference (Berger, 1980). The Bayesian statistician routinely conditionalizes on evidence that outstrips her current mental states. For example, a statistical model might dictate how to update a probability distribution over an external world variable (e.g. the age of a fossil) based on the value of another external world variable (e.g. a carbon dating measurement). Determining the second variable's value is invariably a fallible undertaking.

- *Economics* posits idealized Bayesian agents who conditionalize on propositions that concern the external world. For example, game theory posits Bayesian agents who update credences over properties of other agents after observing how those agents act (Fudenberg and Tirole, 1991, p. 325). One may invest misplaced certainty in propositions about the actions of other agents.

Misplaced certainty arises in virtually every domain where Bayesian updating has found fruitful application: philosophy, economics, statistics, robotics, medicine, cognitive science, and so on.

In each domain, we need a norm that covers misplaced certainty. Philosophers of science need a norm that governs how to update credences based on fallible observational or experimental data. Bayesian statisticians need a norm that governs how to update a probability distribution over one variable based upon fallible determination of another variable's value. Economists need a norm that governs how to update credences based on potentially false beliefs about how other agents acted. Factive formulations of Conditionalization do not even try to cover these and many other theoretically important scenarios. Whatever the merits of factive Conditionalization for certain purposes, it does not provide an adequate foundation for *all* philosophical and scientific applications of Bayesian decision theory.[2]

§3. Jeffrey Conditionalization

[2] The possibility of misplaced certainty bears upon some well-known arguments that seek to justify credal norms. The familiar Dutch Book arguments for Conditionalization and Reflection --- as presented, say, by Skyrms (1987) --- assume that the agent becomes newly certain of a *true* proposition. Furthermore, Schoenfield (2017) shows that expected accuracy arguments for Conditionalization go through only if we assume that the agent is certain she will become newly certain of a *true* proposition. My discussion suggests that these arguments are problematic. I will discuss the implications of misplaced certainty for Dutch book arguments and expected accuracy arguments in future work.

Some readers may respond that scenarios involving misplaced certainty are not so important, since experience rarely authorizes us to invest credence 1 in an empirical proposition. As I just emphasized, evidentiary input to Bayesian inference is almost always fallible. But then shouldn't a rational agent almost always decline to set $P_{new}(E) = 1$? Doesn't the traditional Bayesian emphasis on Conditionalization reflect an outmoded dogmatist epistemology?

Jeffrey (1983, 2004) pursues an alternative approach that does not require new certainties. He focuses on scenarios where experience leads you to reallocate credence across certain select propositions, with no single proposition receiving credence 1. Probability mass shifts without all shifting to a single evidentiary proposition E. To illustrate, suppose you begin with credences given by $P_{old}(\ .\ )$, and suppose you then observe a vase in dim lighting that alters your credences over the propositions:

E_1 = The vase is blue.

E_2 = The vase is green.

E_3 = The vase is neither blue nor green.

to new values $P_{new}(E_i)$, $i = 1, 2, 3$, with $P_{new}(E_i) < 1$. How should you reallocate credence over other propositions (e.g. the proposition *The vase is green & The coin landed heads*)?

Jeffrey (1983) advocates a norm --- now usually called *Jeffrey Conditionalization* --- tailored to learning scenarios of this kind. Earman (1992, p. 34) formulates Jeffrey Conditionalization as follows:

> [L]et $\{E_i\}$, i=1, 2, …, be a partition of the probability space. Intuitively, the belief change that takes place is supposed to be generated by the way in which experience bears on this partition… The belief change then accords with Jeffrey conditionalization just in case
>
> $\mathrm{Pr}_{\mathrm{new}}(A) = \sum_i \mathrm{Pr}_{\mathrm{old}}(A/E_i) \times \mathrm{Pr}_{\mathrm{new}}(E_i)$, for all A.

Thus, Jeffrey conditionalization fixes a unique new credal allocation

(4) $$P_{new}(\,.\,) = \sum_i P_{old}(\,.\,|\,E_i)P_{new}(E_i)\,,$$

for the specified scenario. (2) follows from (4) in the special case where there exists k such that $P_{new}(E_k)$ =1 and $P_{new}(E_i) = 0$ for all $i \neq k$.

Some readers may insist that Jeffrey Conditionalization rather than ordinary Conditionalization is most relevant to typical real-world applications. From this viewpoint, my focus on non-factive Conditionalization is a distraction from the more important task of grounding Bayesianism in Jeffrey Conditionalization.

To a certain extent, I sympathize with these worries. Scenarios involving new certainties are often quite idealized. There is value in studying less idealized scenarios, including Jeffrey-style scenarios. But there is *also* value in studying idealized scenarios that involve new certainties. In field after field --- from statistics to economics to robotics to cognitive science --- researchers have found it explanatorily or pragmatically fruitful to consider such scenarios. As I argued in §2, these scientific applications encompass cases where the conditioning proposition is true along with cases where it is false. Thus, my focus on non-factive Conditionalization is not a distraction from real-world applications. Current scientific applications routinely address updating based upon new certainties, where the certainties may be misplaced.

In any event, we will see that my canonical formulation of Conditionalization generalizes to yield a canonical formulation of Jeffrey Conditionalization. Readers who prefer Jeffrey Conditionalization to ordinary Conditionalization may therefore rest assured that my discussion addresses their preferred model of credal reallocation.

§4. Conditionalization formulated non-factively

A satisfactory non-factive formulation of Conditionalization is more elusive than one might initially expect. There are putative counterexamples to Conditionalization that existing treatments, whether factive or non-factive, do not handle adequately. This section presents a simple counterexample and introduces my preferred solution. Subsequent sections develop the solution in more detail and compare it with alternative solutions.

Consider a problematic formulation of Conditionalization that I do *not* endorse. On the problematic formulation, Conditionalization requires that

(5) If you begin with credences P_{old} such that $0 < P_{old}(E) < 1$, and you subsequently adopt new credences P_{new} such that E is the strongest proposition with $P_{new}(E) = 1$, then $P_{new}(\,.\,) = P_{old}(\,.\, | E)$.

As Jeffrey (2004, p. 52) emphasizes, formulations such as (5) mishandle Jeffrey-style shifts in probability mass. To see why, consider three mutually exclusive, jointly exhaustive propositions E_1, E_2, and E_3 that partition the probability space. You assign these propositions initial credences

$$P_{old}(E_1) = 1/4 \qquad P_{old}(E_2) = 1/4 \qquad P_{old}(E_3) = 1/2.$$

Take a proposition A such that

$$P_{old}(A \mid E_1) = 0 \qquad P_{old}(A \mid E_2) = 1/2 \qquad P_{old}(A \mid E_3) = 0.$$

The probability calculus axioms require that:

$$P_{old}(A \mid E_1 \vee E_2) = 1/4,$$

and you comply. Suppose you gain evidence that strongly favors E_2 but leaves you with a slight degree of belief in E_1. Gaining this evidence leads you to set:

$$P_{new}(E_1) = 1/100 \qquad P_{new}(E_2) = 99/100\ .$$

In conformity with the probability calculus axioms, you set:

$$P_{new}(E_3) = 0 \qquad P_{new}(E_1 \vee E_2) = 1.$$

$E_1 \vee E_2$ is the strongest proposition that acquires credence 1, so the only way to satisfy (5) is to conditionalize on $E_1 \vee E_2$ by setting

$$P_{new}(A) = P_{old}(A \mid E_1 \vee E_2) = 1/4.$$

Intuitively speaking, though, $P_{new}(A) = 1/4$ seems quite inappropriate. Probability mass has shifted markedly from E_1 to E_2, so it looks like the new credence for A should be much higher. The credence given by Jeffrey conditionalization

$$P_{new}(A) = P_{old}(A \mid E_1)\, P_{new}(E_1) + P_{old}(A \mid E_2)\, P_{new}(E_2) + P_{old}(A \mid E_3)\, P_{new}(E_3)$$

$$= 0 \times \tfrac{1}{100} + \tfrac{1}{2} \times \tfrac{99}{100} + 0 \times 0 = \tfrac{99}{200}\ .$$

looks plausible. Whether or not we accept the answer given by Jeffrey conditionalization, conditionalizing on $E_1 \vee E_2$ is plainly not a good idea. Indeed, conditionalizing on $E_1 \vee E_2$ would lead you to assign credence ½ to E_1 and to E_2, which is incompatible with the postulated shift in probability mass from E_1 to E_2. Thus, (5) faces a devastating Jeffrey-style counterexample.

INSERT FIGURE 1 ABOUT HERE

The problem posed by Jeffrey-style counterexamples is not peculiar to (5). It impacts most formulations of Conditionalization found in the current literature, factive and non-factive alike. Consider Earman's factive formulation, quoted in §1. Suppose you "learn for sure" that $E_1 \vee E_2$ and *simultaneously* shift probability mass from E_1 to E_2. For example, you "learn for sure" that the vase is either blue or green ($E_1 \vee E_2$) while *simultaneously* becoming much more confident but not completely confident that it is green (E_2). You should not update your credences by conditionalizing on $E_1 \vee E_2$, yet Earman's formulation requires you to do so.

Perhaps what Earman has in mind is not simply that you "learn E for sure" but also that all credal changes result *solely* from the fact that you "learn E for sure." Yet that is not what Earman says.

Missing from Earman's formulation and virtually all other formulations in the literature is any explicit acknowledgment of the distinctive *causal* role played by new certainties. The reason why you should not conditionalize in the Jeffrey-style counterexample is that your newfound certainty in $E_1 \vee E_2$ is not the sole source of change in your credences. There is also a shift in probability mass from E_1 to E_2, and this shift is not mediated by your new certainty in $E_1 \vee E_2$.

I propose that we incorporate more explicitly causal language into our formulation of Conditionalization. Here is an initial attempt at formulating what Conditionalization requires:

(6) If you begin with credences P_{old} such that $0 < P_{old}(E) < 1$, and you subsequently adopt new credences P_{new} such that $P_{new}(E) = 1$, and the new credal assignment to E mediates the transition from P_{old} to P_{new}, then $P_{new}(\,.\,) = P_{old}(\,.\, | E)$.

Intuitively, (6) restricts attention to situations where your credences change entirely as a result of your newfound certainty in E.[3]

Say that a *credal reallocation event* is an event that changes your credences. A credal reallocation event may influence your credences through perception (e.g. you see a red cube), sensation (e.g. you feel a sharp pain in your knee), testimony (e.g. your history teacher tells you when the Civil War ended), episodic memory (e.g. you remember breaking your leg when you were a child), introspection (e.g. you realize that you have a craving for ice cream), *a priori* reasoning (e.g. you construct a new *a priori* argument for the principle "Every event has a

[3] (5) incorporates a "total evidence" requirement: E is the strongest proposition with $P_{new}(E) = 1$. This requirement is the usual way that authors specify a unique privileged conditioning proposition (up to logical equivalence) among all those propositions that receive new credence 1. The total evidence requirement favors E rather than some weaker logical consequence of E as the conditioning proposition. (6) incorporates no total evidence requirement. Instead, (6) specifies a unique privileged conditioning proposition by invoking the causal structure of credal evolution. If the new credal assignment to E mediates the transition from P_{old} to P_{new}, then no distinct proposition plays that same mediating role. (6) therefore pinpoints E rather than any weaker logical consequence as the appropriate conditioning proposition.

cause”), or other possible epistemic channels. A credal reallocation event may also exert *non-epistemic* influence upon your credences (e.g. a bump on the head causes you to believe that you are an architect; or you ingest a mind-altering pill with the same effect). There are many possible credal reallocation events, and those events can exert diverse modes of causal influence upon your credences. (6) restricts attention to scenarios where credal allocation events influence P_{new} entirely by instilling certainty in proposition E. The only *direct* influence that any credal reallocation event exerts on P_{new} is fixing $P_{new}(E) = 1$. When the transition from P_{old} to P_{new} occurs thus, the only way to satisfy (6) is to conditionalize on E.

(6) blocks our earlier Jeffrey-style counterexample. The counterexample features a credal reallocation event that shifts probability mass from E_1 to E_2. The shift is not mediated by your setting $P_{new}(E_1 \vee E_2) = 1$. The credal reallocation event most directly affects credences assigned to E_1 and E_2, not the credence assigned to $E_1 \vee E_2$. So the transition from P_{old} to P_{new} is not mediated by the altered credal assignment to $E_1 \vee E_2$ or to any other proposition that receives credence 1, and (6) issues no instruction on how to proceed.

Although (6) improves upon our previous efforts, it does not cover certain cases we would like to cover. Take the same propositions E_1, E_2, E_3, and A with the same initial unconditional and conditional probabilities from our Jeffrey-style counterexample. Suppose you gain overwhelming evidence against E_1 and separate overwhelming evidence against E_3. This leads you to set $P_{new}(E_1) = 0$ and $P_{new}(E_3) = 0$, which in turns leads you to set $P_{new}(E_2) = 1$. (6) is trivially satisfied no matter what new credence you assign to A, because the transition to $P_{new}(E_1)$ and $P_{new}(E_3)$ is not mediated by the new credal assignment to E_2. Thus, (6) issues no instruction regarding your new credence in A.

Ideally, we would like a formulation that covers this sort of case. I propose that we revise (6) as follows. Let $\boldsymbol{E} = \{E_i\}$ be a countable set of mutually exclusive, jointly exhaustive propositions. These propositions partition the probability space. I restrict attention to situations where each credal reallocation event *c* influences credal assignments to propositions $H \notin \boldsymbol{E}$ solely by influencing credal assignments to propositions in $\boldsymbol{E}$. It may be that *c* directly fixes individual credal assignments for each proposition in $\boldsymbol{E}$, or it may be that *c* directly fixes new credal assignments for *certain* propositions in $\boldsymbol{E}$ and that these new credal assignments fix new credal assignments for remaining members of $\boldsymbol{E}$. What matters is that *c* influences credal assignments to propositions *outside* $\boldsymbol{E}$ only by influencing credal assignments to propositions *inside* $\boldsymbol{E}$. Conditionalization requires that:

COND	If you begin with credences P_{old}, and $\boldsymbol{E} = \{E_i\}$ is a countable set of mutually exclusive, jointly exhaustive propositions such that $P_{old}(E_i) > 0$ for each *i*, and you subsequently adopt new credences P_{new} such that $P_{new}(E_k) = 1$ and $P_{new}(E_i) = 0$ for each $i \neq k$, and the new credal assignment over $\boldsymbol{E}$ mediates the transition from P_{old} to P_{new}, then $P_{new}(\ .\ ) = P_{old}(\ .\ \vert\ E_k)$.

Intuitively, we restrict attention to situations where there is an exogenous shift in probability mass to a single proposition E_k from $\boldsymbol{E}$. COND handles both our original Jeffrey-style counterexample *and* the modified example from the preceding paragraph.

Another virtue of COND is that it naturally generalizes to yield a kindred formulation of what Jeffrey Conditionalization requires:

JCOND If you begin with credences P_{old}, and $\boldsymbol{E} = \{E_i\}$ is a countable set of mutually exclusive, jointly exhaustive propositions such that $P_{old}(E_i) > 0$ for each i, and you subsequently adopt new credences P_{new} such that $\sum_i P_{new}(E_i) = 1$, and the new credal assignment over $\boldsymbol{E}$ mediates the transition from P_{old} to P_{new}, then $P_{new}(\,.\,) = \sum_i P_{old}(\,.\,|\,E_i)P_{new}(E_i)$.

Intuitively, we restrict attention to situations where there is an exogenous shift in probability mass across the partition $\boldsymbol{E}$. COND follows from JCOND in the special case where there exists k such that $P_{new}(E_k) = 1$ and $P_{new}(E_i) = 0$ for all $i \neq k$.

§5. The mediation clause

My formulation COND features the following clause in the antecedent: *the new credal assignment over* $\boldsymbol{E}$ *mediates the transition from* P_{old} *to* P_{new}. Call this *the mediation clause*. I now elucidate the mediation clause, using tools from the recent literature on causation.

§5.1 Causal structure

Philosophers standardly distinguish between *type causal claims* and *token causal claims*, as exemplified by the contrast between the following two statements:

Smoking causes cancer.

Jones's smoking caused his cancer.

The relation between type causation and token causation is complex. For example, it may be that smoking causes cancer, that John smokes, that John contracts cancer, yet that John's smoking did not cause his cancer. In what follows, I will not assume any specific theory of type causation, token causation, or the relation between them.

It is increasingly common to regiment causal claims using *variables* (Woodward, 2003). The *values* of a variable are possible states of the system under consideration. Any variable has at least two values. Some variables have infinitely many values. Following Woodward, I regiment type causal claims using the schema:

(7) X is causally relevant to Y,

where X and Y are variables. Roughly, (7) means that which value X assumes makes a difference to which value Y assumes. For example, we can introduce a binary variable X whose two values reflect whether John smokes and a binary variable Y whose two values reflect whether John contracts cancer. We can then say that X is causally relevant to Y, meaning that:

Whether John smokes is causally relevant to whether John contracts cancer.

Note that this regimentation treats causal relevance as *contrastive*. Type-causal relations obtain between variables, which enshrine contrast classes of possible values. There are strong reasons, independent of any specific theory of causation, for treating causal relevance as contrastive (Schaffer, 2005; Woodward, 2003, pp. 145-146). Again following Woodward, I regiment token causal claims as follows:

(8) $X = x$ is an actual cause of $Y = y$,

e.g. John's smoking was an actual cause of his contracting cancer. Any complete theory of causation must elucidate (7) and (8) in great detail. The literature offers several compelling treatments that yield intuitively correct verdicts across a wide range of cases (Halpern and Hitchcock, 2015; Pearl, 2000; Woodward, 2003). I will take (7) and (8) as primitive.

I also presuppose a notion of *direct causation.* Let V be a set of variables. Roughly, X is *directly causally relevant to Y relative to V* iff X is causally relevant to *Y and X* influences Y through a causal route that passes through no other variables in V. To use Woodward's example, taking the birth control pill raises the probability of thrombosis. However, taking the birth control pill also lowers the probability of getting pregnant, and getting pregnant raises the probability of thrombosis. Intuitively, there are two causal routes through which taking the pill influences whether thrombosis occurs. Figure 2 depicts these two routes. Figure 2 contains three variables: B (a binary variable whose value reflects whether the agent takes the birth control pill), P (a binary variable whose value reflects whether she becomes pregnant), and T (a binary variable whose value reflects whether thrombosis occurs). One causal route goes from B to T through P. The second goes directly from B to T. Thus, B is directly causally relevant to T relative to $V = \{B, P, T\}$. For one possible elucidation of *direct causation*, see (Woodward, 2003, p. 55). I leave the notion at an intuitive level.

INSERT FIGURE 2 ABOUT HERE

I adopt the following conventions regarding causal diagrams such as Figure 2: an arrow from X to Y means that X is causally relevant to Y via a causal route that does not go through any other variable in the causal diagram; no arrow from X to Y means that any causal route from X to Y goes through other variables in the causal diagram; no sequence of arrows leading from X to Y means that X is not causally relevant to Y. For instance, Figure 3 entails that

> X is causally relevant to Y.
>
> Y is causally relevant to Z.
>
> X is not directly causally relevant to Z relative to $\{W, X, Y, Z\}$
>
> W is not causally relevant to X, Y, or Z.

Figure 3 is neutral about whether X is causally relevant to Z. Thus, I do not assume transitivity of causal relevance.

INSERT FIGURE 3 ABOUT HERE

Say that variables $X_1, X_2, \ldots, X_n$ are *independently manipulable* iff it is metaphysically possible that

$$X_1 = x_1 \;\&\; X_2 = x_2 \;\&\; \ldots \;\&\; X_n = x_n$$

for all individually possible values $x_1, x_2, \ldots, x_n$ of the respective variables. For example, two variables are not independently manipulable if their values co-vary due to logic, definition, or supervenience. It is common to require that variables figuring in the same causal diagram be independently manipulable, so that distinct elements of the diagram represent causal factors that are totally distinct from one another. There are some occasions, such as when studying the interplay between causation and supervenience, when one may want to consider variables that are *not* independently manipulable (Woodward, 2015). However, independently manipulable variables suffice for most modeling purposes.

The final notion I will presuppose is *intervention on X with respect to Y*. Intuitively, an intervention is an exogenous manipulation of X that changes Y, if at all, only by changing X. A paradigmatic example is an experimental manipulation designed to test whether X is causally relevant to Y. For instance, we might test whether cholesterol levels are relevant to heart disease by administering a drug that lowers cholesterol levels and that does not alter any confounding variables. Interventions figure crucially in much of the recent causation literature.

The literature offers various subtly different definitions of "intervention." Woodward's (2003, pp. 98-99) definition has been especially influential. The definition of "mediation" that I give below could be combined with different definitions of "intervention," yielding the same results except for quite exotic cases. Hoping to remain as neutral as possible, I take as a primitive the locution:

(9) $C = c$ is an intervention on X with respect to Y.

I assume that (9) has several consequences:

(I.1) C is causally relevant to X.

(I.2) $C = c$ is an actual cause of $X = x$, for some x.

(I.3) If C and W are independently manipulable, and C is directly causally relevant to X relative to $\{C, W, X\}$, then W is not causally relevant to X.

(I.4) C is not directly causally relevant to Y relative to $\{C, X, Y\}$.

Clauses I.1-I.4 are necessary but not sufficient for $C = c$ to be an intervention X with respect to Y. Clauses I.1-I.3 reflect the intuitive thought that C having value c "breaks" all other causal routes into X, pinning X to value x. Figures 4 and 5 illustrate: the effect of the intervention $C = c$ is to replace the first causal diagram, in which W is causally relevant to X, with a second diagram in which W is no longer causally relevant to X. When clauses I.1-I.3 hold, I say that

$C = c$ controls X.

Clause I.4 reflects the intuitive thought that any causal route from C to Y goes through X.[4] Clause 1.4 is satisfied by Figure 5 but not by Figure 6. Clauses I.1-I.4, or clauses much like them, are standard in the literature. Although human manipulations are paradigmatic examples of interventions, there is nothing anthropomorphic about clauses I.1-I.4 or about interventions more generally. Interventions may arise independently of any human action, as in "natural experiments."

INSERT FIGURES 4-6 ABOUT HERE

Woodward (2003) deploys the notion of *intervention* to elucidate causal relevance. My discussion here does not presuppose this elucidatory strategy. I assume that interventions are useful tools for studying causal structure, but I do not assume that we should invoke them so as to elucidate causal relevance itself.[5]

§5.2 The causal structure of credal reallocation

Consider an agent who begins with credences P_{old} and transitions to new credences P_{new}. In principle, there are many possible causal structures that might underwrite the transition. I am concerned with situations where the transition is sparked entirely by an exogenous change in the credal assignment over propositions belonging to some set $\boldsymbol{E}$. This exogenous change "breaks"

[4] To capture the intuitive thought more fully, one must strengthen I.4 so as to prohibit any sequence of arrows from C to Y that bypasses X. This is one reason why I.1-I.4 are necessary but not sufficient for $C = c$ to be an intervention on X with respect to Y.

[5] In (Rescorla, 2014), I argue that interventions shed light upon mental causation.

whatever influence P_{old} would normally exert upon the agent's new credences over $\boldsymbol{E}$, and it influences her credences outside $\boldsymbol{E}$ solely by influencing her credences over $\boldsymbol{E}$.

The notion of *intervention* is well-suited to capture these intuitive ideas. We may regard P_{old} as a variable whose value reflects the initial credal allocation. Let $\boldsymbol{H}$ and $\boldsymbol{E}$ be disjoint sets of propositions, and let $P_{new}(\boldsymbol{H})$ and $P_{new}(\boldsymbol{E})$ be variables that reflect the new credal assignments over $\boldsymbol{H}$ and $\boldsymbol{E}$ respectively. I am concerned with situations where some intervention affects $P_{new}(\boldsymbol{H})$, if at all, only through its impact on $P_{new}(\boldsymbol{E})$. More precisely:

(10) There exist a variable C and a value c such that C and P_{old} are independently manipulable and such that, for every $\boldsymbol{H}$ that is disjoint from $\boldsymbol{E}$, $C = c$ is an intervention on $P_{new}(\boldsymbol{E})$ with respect to $P_{new}(\boldsymbol{H})$.

For any such C and c, we have:

$C = c$ controls $P_{new}(\boldsymbol{E})$.

C is not directly causally relevant to $P_{new}(\boldsymbol{H})$ relative to $\{C, P_{new}(\boldsymbol{E}), P_{new}(\boldsymbol{H})\}$.

See Figure 7.

INSERT FIGURE 7 ABOUT HERE

(10) describes an abstract causal structure that can be instantiated in diverse ways by diverse credal reallocation scenarios. In some cases, we may take C to be a binary variable whose two values reflect whether some credal reallocation event occurs. In other cases, we may want to consider a variable C that has many possible values. For example, C's values may reflect possible visual experiences you might undergo. There are sometimes multiple causal factors that simultaneously affect credences over $\boldsymbol{E}$: a visual experience; a haptic experience; your relative

inclination to take visual versus haptic experience at face value; how carefully you are paying attention; and so on. We may assimilate these multiple causal factors into a single variable C that is causally relevant to $P_{new}(\boldsymbol{E})$.

To see (10) in action, suppose that a credal reallocation event renders you certain of E and that this is the only exogenous influence on your credences. Suppose that you respond to your newfound certainty in E by conditionalizing on E, using (3) to compute new credences. Let C be a binary variable whose two values reflect whether the credal reallocation event occurs: 1 signifies that it occurs; 0 signifies that it does not. Quite plausibly, $C = 1$ is an intervention on $P_{new}(\{E, \neg E\})$ with respect to $P_{new}(\boldsymbol{H})$, for any set of propositions $\boldsymbol{H}$ disjoint from $\{E, \neg E\}$. Thus, Figure 8 gives a plausible representation of the scenario's causal structure.[6] Scenarios where you Jeffrey conditionalize are handled similarly, except that in those scenarios there is an intervention on $P_{new}(\{E_i\})$ with respect to $P_{new}(\boldsymbol{H})$ for some partition $\{E_i\}$. See Figure 9. Finally, consider the Jeffrey-style counterexample from §4. There is an intervention on $P_{new}(\{E_1, E_2, E_3\})$ with respect to $P_{new}(\boldsymbol{H})$, where $\boldsymbol{H}$ is any set of propositions disjoint from $\{E_1, E_2, E_3\}$. Specifically, there is an intervention on $P_{new}(\{E_1, E_2, E_3\})$ with respect to $P_{new}(E_1 \vee E_2)$. In contrast, there is no intervention on $P_{new}(\{E_1 \vee E_2, E_3\})$ with respect to $P_{new}(E_1)$ or $P_{new}(E_2)$. The exogenous influence most directly alters $P_{new}(\{E_1, E_2, E_3\})$ rather than $P_{new}(\{E_1 \vee E_2, E_3\})$. See Figure 10.

INSERT FIGURES 8-10 ABOUT HERE

[6] One can imagine an agent who automatically assigns credence 1 to any logical truth, without bothering to recompute that credence in light of new evidence. For such an agent, $P_{new}(\{E, \neg E\})$ is not causally relevant to $P_{new}(\boldsymbol{H})$ when $\boldsymbol{H}$ contains only logical truths. This alternative causal structure is also consistent with (10).

Although (10) is a good start towards defining mediation, it is too weak when taken on its own because it falls prey to a modified Jeffrey-style counterexample. Suppose that there are two distinct, simultaneous credal reallocation events:

- The first event renders you certain of $E_1 \vee E_2$ --- e.g. a guru whom you completely trust tells you that the vase is either green or blue.
- The second event leads to a shift in credal mass over $\{E_1, E_2\}$ --- e.g. visual observation of the vase leads you to set $P_{new}(E_1) = {}^{1}\!/_{100}$ and $P_{new}(E_2) = {}^{99}\!/_{100}$.

Let C be a binary variable that reflects whether the first event occurs. Let D be a binary variable that reflects whether the second occurs. For each variable, 1 signifies that the relevant event occurs and 0 signifies that it does not. Let $\boldsymbol{H}$ be any set of propositions disjoint from both $\{E_1 \vee E_2, E_3\}$ and $\{E_1, E_2\}$. We may stipulate that:

- $C = 1$ is an intervention on $P_{new}(\{E_1 \vee E_2, E_3\})$ with respect to $P_{new}(\boldsymbol{H})$.
- $D = 1$ is an intervention on $P_{new}(\{E_1, E_2\})$ with respect to $P_{new}(\boldsymbol{H})$.

Figure 11 depicts the causal structure. Given the credal shift over $\{E_1, E_2\}$, it hardly seems like a good idea to conditionalize on $E_1 \vee E_2$. Yet that is what COND would demand were we to take (10) as our definition of mediation.[7]

INSERT FIGURE 11 ABOUT HERE

Clause (10) on its own does not fully capture our intuitive starting point. We want to consider scenarios where the credal transition is sparked *entirely* by an exogenous change in

[7] In the modified Jeffrey counterexample, $D = 1$ controls $P_{new}(\{E_1, E_2\})$, so Figure 11 contains no arrow into $P_{new}(\{E_1, E_2\})$ from any variable other than D. One can also construct counterexamples in which $D = 1$ does not control $P_{new}(\{E_1, E_2\})$ but is still an actual cause of $P_{new}(\{E_1, E_2\})$ having the value it has. For reasons of space, I will not explore such counterexamples here.

credences over $\boldsymbol{E}$. It is not enough to require that there be an intervention on $P_{new}(\boldsymbol{E})$. We must furthermore require that there be *no additional influence* besides P_{old} and $P_{new}(\boldsymbol{E})$ on credences outside $\boldsymbol{E}$. All causal influence on $P_{new}(\boldsymbol{H})$ should run through either P_{old} or $P_{new}(\boldsymbol{E})$.

Formulating the "no additional influence" requirement is a bit tricky. Whenever we draw a causal diagram, we omit intermediate variables that do not bear on our current concerns. For example, $P_{new}(\boldsymbol{E})$ may impact $P_{new}(\boldsymbol{H})$ via intermediary mental computations. Figure 12 depicts two intermediate variables D_1 and D_2. Depending on the details, $D_1 = d_1$ or $D_2 = d_2$ may be an actual cause of $P_{new}(\boldsymbol{H}) = \alpha$. We do not want to rule out such "intermediate causes." We only want to rule out additional causes that fall outside any route emanating from P_{old} or $P_{new}(\boldsymbol{E})$. More precisely, say that a sequence $\langle X_1, \ldots, X_n \rangle$ is a *directed path from* X_1 *to* X_n iff X_i is directly causally relevant to X_{i+1} relative to $\{X_1, \ldots, X_n\}$, for each $i<n$. Say that *variable* Z *falls outside directed path* $\langle X_1, \ldots, X_n \rangle$ iff $Z \neq X_i$ for any i. We want to rule out additional causes $D = d$ such that D falls outside every directed path from $P_{new}(\boldsymbol{E})$ to $P_{new}(\boldsymbol{H})$ and every directed path from P_{old} to $P_{new}(\boldsymbol{H})$.

INSERT FIGURE 12 ABOUT HERE

I now define "mediation" in a way that incorporates these ideas. Say that *the new credal assignment over* $\boldsymbol{E}$ *mediates the transition from* P_{old} *to* P_{new} iff there exist C and c such that C and P_{old} are independently manipulable and such that

(M.1) For every $\boldsymbol{H}$ that is disjoint from $\boldsymbol{E}$, $C = c$ is an intervention on $P_{new}(\boldsymbol{E})$ with respect to $P_{new}(\boldsymbol{H})$.

(M.2) There exist no $\boldsymbol{H}$, α, D, and d such that

a. $\boldsymbol{H}$ is disjoint from $\boldsymbol{E}$.

b. C, D, P_{old}, $P_{new}(\boldsymbol{E})$, and $P_{new}(\boldsymbol{H})$ are independently manipulable.

c. D is directly causally relevant to $P_{new}(\boldsymbol{H})$ relative to $\{C, D, P_{old}, P_{new}(\boldsymbol{E}), P_{new}(\boldsymbol{H})\}$.

d. D falls outside every directed path from $P_{new}(\boldsymbol{E})$ to $P_{new}(\boldsymbol{H})$.

e. D falls outside every directed path from P_{old} to $P_{new}(\boldsymbol{H})$.

f. $D = d$ is an actual cause of $P_{new}(\boldsymbol{H}) = \alpha$.

See Figures 13 and 14. Roughly speaking, M.2 serves to prohibit the causal structures depicted in Figures 13 and 14 while allowing the causal structure depicted in Figure 12.

INSERT FIGURES 13 AND 14 ABOUT HERE

My definition of "mediation" invokes token causation, so it yields different results when combined with different theories of token causation. Some theorists, such as Hall (2004) and Lewis (1986), hold that a typical event has many actual causes. To illustrate, suppose that lightning strikes a house and starts a fire. Everyone agrees that the lighting causes the fire. Hall maintains that there are many additional causes, including that oxygen is in the air, that flammable material is present inside the house, and so on. Similarly, one might insist that $P_{new}(\boldsymbol{H}) = \alpha$ has numerous causes: that there is sufficient blood flow to your brain; that you have not suffered a concussion; that you have not taken a pill fixing $P_{new}(\boldsymbol{H})$ to value $\beta \neq \alpha$; and so on. Any such proliferation of causes would devastate my account, since it would entail that all credal reallocation scenarios instantiate Figure 13 and that clause M.2 is never satisfied.

Luckily, we need not proliferate causes in the manner of Hall and Lewis. Many attractive, well-developed theories of token causation are vastly more selective in what they count as actual causes (e.g. Halpern and Hitchcock, 2015; Woodward, 2003, pp. 74-91). Selective theorists can say that whether there is oxygen in the air is causally relevant to whether the house catches fire *and yet* that the presence of oxygen is not an actual cause of the fire; that whether you take a credence altering pill is causally relevant to $P_{new}(\boldsymbol{H})$ *and yet* that your failure to take the pill is not an actual cause of $P_{new}(\boldsymbol{H}) = \alpha$; and so on. Selective theorists differ as to *why*, in the cases under consideration, $X = x$ is not an actual cause of $Y = y$ even though X is causally relevant to Y. The basic idea behind most treatments is that $X = x$ counts as a background condition or default state, rather than an actual cause. The selective approach fits much better with practice both in ordinary life and in various systematic endeavors where causation plays a role, such as legal reasoning (Halpern and Hitchcock, 2015). Thus, there are strong reasons quite independent of my discussion for favoring the selective approach.

If we adopt the selective approach, then we can say that clause M.2 is satisfied in a huge range of scenarios. Not *all* scenarios, of course, but all or at least virtually all scenarios where mediation by $\boldsymbol{E}$ intuitively takes place. For example, the selective approach combined with my definition of "mediation" supports the following verdicts:

- In the scenario corresponding to Figure 8, the new credal assignment over $\{E, \neg E\}$ mediates the credal transition.
- In the scenario corresponding to Figure 9, the new credal assignment over $\{E_i\}$ mediates the credal transition.
- In the scenario corresponding to Figure 10, the new credal assignment over $\{E_1, E_2, E_3\}$ mediates the credal transition but the new credal assignment over $\{E_1 \vee E_2, E_3\}$ does not.

- In the scenario corresponding to Figure 11, there is no set $\boldsymbol{E}$ of mutually exclusive, jointly exhaustive propositions such that the new credal assignment over $\boldsymbol{E}$ mediates the credal transition.

All these verdicts are plausible.[8]

Philosophical discussion over the past few decades has amply demonstrated that token causation is an extraordinarily useful concept, especially in normative contexts such as the assessment of moral and legal responsibility. There are diverse circumstances, both in everyday life and in systematic inquiry, where normative evaluation hinges upon verdicts regarding what caused what. My analysis highlights one further circumstance: normative evaluation of credal reallocation. By elucidating "mediation" in causal terms, we clarify the fulfillment-conditions for Conditionalization and Jeffrey Conditionalization.

I offer my elucidation in a preliminary spirit. I encourage readers to explore alternative elucidations, which may well prove superior.

§6. Alternatives to the mediation clause

My preferred formulation COND may seem to veer too far into speculative psychology, encouraging ill-supported conjectures about the causal structure of credal reallocation. Wouldn't it be better to formulate Conditionalization without citing anything like the mediation clause? Do

[8] Theorists who reject the selective approach may consider revising my definition of "mediation" in an effort to achieve similar results. One idea is to supplement M.2 with a further clause M.2.g along the following lines: *If the causal link between D and* $P_{new}(\boldsymbol{H})$ *had been broken, so that D was no longer causally relevant to* $P_{new}(\boldsymbol{H})$*, and if* P_{old} *and* $P_{new}(\boldsymbol{E})$ *had had the same values as they in fact do, then* $P_{new}(\boldsymbol{H})$ *would not have had value* α. The effect of M.2.g is to allow actual causes that make no real difference to $P_{new}(\boldsymbol{H})$. To illustrate, let D be a binary variable whose two values reflect whether you take a pill that sets $P_{new}(\boldsymbol{H}) = \beta \neq \alpha$ (1 signifies that you take the pill; 0 signifies that you do not). Suppose that $D = 0$. Then D violates M.2.g: you could break D's causal influence on $P_{new}(\boldsymbol{H})$ by ingesting a second pill that counteracts the credence-altering properties of the first pill, and doing so would not change $P_{new}(\boldsymbol{H})$ *because you did not in fact ingest the first pill*. Thus, even if we count $D = 0$ as an actual cause of $P_{new}(\boldsymbol{H}) = \alpha$, the revised definition of "mediation" still allows us to say that the new credal assignment over $\boldsymbol{E}$ mediates the credal transition. I leave further exploration of such revisionary stratagems to those who reject the selective approach.

Jeffrey-style shifts really force us to pollute our account with causal notions? I now address these concerns.

§6.1 Invariant conditional probabilities

The existing literature is surprisingly little help here. On the one hand, few if any discussants would apply Conditionalization to situations where Jeffrey-style shifts occur. On the other hand, authors seldom explicitly address how one should formulate Conditionalization so as to forestall such misapplications.

Jeffrey (2004, pp. 51-52) offers one of the few systematic treatments. He proposes that we restrict Conditionalization to situations where probabilities conditional on E are *invariant*:

(11) $P_{old}(H \mid E) = P_{new}(H \mid E)$ for all propositions H.

In this vein, one might formulate Conditionalization as requiring that:

(12) If you begin with credences P_{old} such that $0 < P_{old}(E) < 1$, and you subsequently adopt new credences P_{new} such that E is the strongest proposition with $P_{new}(E) = 1$, and $P_{old}(\,.\, \mid E) = P_{new}(\,.\, \mid E)$, then $P_{new}(\,.\,) = P_{old}(\,.\, \mid E)$.

In our initial Jeffrey-style counterexample, someone who updates using Jeffrey Conditionalization will set

$$P_{old}(A \mid E_1 \vee E_2) \neq P_{new}(A \mid E_1 \vee E_2),$$

so (12) does not require such an agent to conditionalize on $E_1 \vee E_2$.

The problem with (12) is that it lacks any non-trivial dynamic element. Assume that P_{new} satisfies the probability calculus axioms. If (12)'s antecedent is true, then for any H

$$P_{new}(H) = P_{new}(H \mid E) = P_{old}(H \mid E),$$

so long as P_{new} conforms to the probability calculus axioms. In other words, it is impossible for an agent whose credences satisfy the probability calculus axioms to violate (12). For this reason, (12) does not impose a non-trivial *diachronic* constraint on the relation between P_{old} and P_{new}. In effect, (12) imposes a purely *synchronic* constraint: namely, that P_{new} conform to the probability calculus axioms. Conditionalization is supposed to require more than that P_{new} conform to the probability calculus axioms.

The problem arises because (12) *assumes* invariant conditional probabilities in the antecedent. We should instead isolate conditions under which invariant conditional probabilities are required. That is precisely what COND accomplishes. If you conditionalize on E, and if P_{new} conforms to the probability calculus axioms, then

$$P_{new}(H \mid E) = P_{new}(H) = P_{old}(H \mid E),$$

so that (11) is satisfied. If you conform to the probability calculus axioms, then the only way for you to satisfy both the antecedent and the consequent of COND is to satisfy (11). Unlike (12), COND imposes a non-trivial diachronic constraint. Even if you satisfy the probability calculus axioms, you can satisfy COND's antecedent while violating its consequent. For example, you may have an irrational policy of shifting probability mass from E_1 to E_2 whenever you set $P_{new}(E_1 \vee E_2) = 1$, even if your evidence provides no good reason to do so. Hence, you can violate COND even if your new credences satisfy the probability calculus axioms. In other words, COND demands more than simply that P_{new} conform to the probability calculus axioms.

To clarify the relation between the Conditionalization and invariant conditional probabilities, consider the following plausible requirement upon credal reallocation:

(13) If you begin with credences P_{old}, and $\boldsymbol{E} = \{E_i\}$ is a countable set of mutually exclusive, jointly exhaustive propositions such that $P_{old}(E_i) > 0$ for each i, and you subsequently adopt new credences P_{new} such that $\sum_i P_{new}(E_i) = 1$, and the new credal assignment over $\boldsymbol{E}$ mediates the transition from P_{old} to P_{new}, then $P_{old}(\,.\, | E_i) = P_{new}(\,.\, | E_i)$ for all i such that $P_{new}(E_i) > 0$.

Intuitively: reallocating probabilities *across* a partition tells you nothing new about how probability mass should be distributed *inside* any given member of the partition. To borrow Joyce's (2009) label for a closely related requirement, (13) is a *minimal change* principle. (13) and the law of total probability jointly entail JCOND, which entails COND. In this manner, (13) helps explain why credal reallocation should satisfy COND.

§6.2 Foregoing mediation?

Building upon §6.1, let us investigate more generally how one might try to avoid the mediation clause. Consider a schematic formulation of what Conditionalization requires:

(14) If you begin with credences P_{old} such that $0 < P_{old}(E) < 1$, and you subsequently adopt new credences P_{new} such that E is the strongest proposition with $P_{new}(E) = 1$, and $\Gamma(P_{old}, P_{new})$, then $P_{new}(\,.\,) = P_{old}(\,.\, | E)$.

To fill in the schema, one chooses a particular relation Γ. Say that Γ is *mathematical* when the truth of $\Gamma(P_{old}, P_{new})$ depends solely upon properties of P_{old} and P_{new} *viewed as functions from propositions to real numbers*. For example, the relation Γ given by

$\Gamma(P_{old}, P_{new})$ iff $P_{old}(\,.\, | E) = P_{new}(\,.\, | E)$

is mathematical. The relation Γ given by

$\Gamma(P_{old}, P_{new})$ iff the new credal assignment over $\boldsymbol{E}$ mediates the transition from your old credences P_{old} to your new credences P_{new}

is not mathematical, because it depends on causal relations between the credal states modeled by functions P_{old} and P_{new}. I will now argue that no mathematical Γ yields a satisfactory formulation (14) of Conditionalization.

Suppose first that (14)'s antecedent entails invariant conditional probabilities (11). Then it is impossible for an agent whose new credences P_{new} satisfy the probability calculus axioms to violate (14). As in §6.1, (14) does not place a substantive dynamic constraint on credal evolution. Thus, a satisfactory choice of Γ must *not* make (14)'s antecedent entail invariant conditional probabilities (11).

Suppose, then, that (14)'s antecedent does not entail (11). I will sketch a three-stage scenario where your credal reallocation looks reasonable even though you satisfy (14)'s antecedent and violate its consequent. For purposes of the scenario, I suppose that the set of relevant propositions is finite --- such situations arise in many applications.

STAGE 1: You begin with credences given by P_{old}.

STAGE 2: There occurs a credal reallocation event whose only immediate effect on your credences is to render you certain of E. You respond by conditionalizing on E, forming new credences given by $P_E =_{df} P_{old}(\,.\, | \, E)$. As a result, E is now the strongest proposition of which you are certain.

Before I describe Stage 3, note that our assumptions on Γ entail the existence of at least one function P_{new} such that

$\Gamma(P_{old}, P_{new})$.

E is the strongest proposition with $P_{new}(E) = 1$.

P_{new} satisfies the probability calculus axioms.

$P_{new} \neq P_{old}(\,.\, | \, E) = P_E$.

Given any such P_{new}, one can prove that there exists a finite set of mutually exclusive, jointly exhaustive propositions $\{E_i\}$ with

(15) $P_{new}(\, . \mid E_i) = P_E(\, . \mid E_i)$ for each i such that $P_{new}(E_i) > 0$.

For a proof, see (Williamson, 2000, pp. 216-217, fn. 2).[9] Now here is Stage 3:

> **STAGE 3**: You undergo a Jeffrey-style shift that exogenously reallocates credence over the partition $\{E_i\}$. The shift causes you to transition from P_E to new credences P_{new}.

Since Γ is mathematical, it only constrains P_{old} and P_{new} *viewed as functions from propositions to real numbers*. So Γ cannot rule out the described three-stage scenario. Your credences P_{old} and P_{new} satisfy (14)'s antecedent. Yet they violate (14)'s consequent. Everyone will agree that Stage 2 is reasonable. Stage 3 also looks reasonable, since P_{new} satisfies the invariance condition (15). Thus, I have sketched a scenario where your overall credal reallocation looks reasonable even while it violates (14). I conclude that (14) is an unreasonable constraint upon credal evolution.

The basic problem here is that we want to rule out Jeffrey-style shifts without simply *assuming* invariant conditional probabilities (11). We cannot accomplish that goal by constraining mathematical properties of P_{old} and P_{new} *viewed as functions from propositions to numbers*. We must instead constrain the *process* whereby you transition from a credal state modeled by P_{old} to a credal state modeled by P_{new}.

What if we employ a non-mathematical Γ? Can we find a non-mathematical constraint that rules out Jeffrey-style shifts while eschewing anything like mediation? I doubt it. I suspect that any adequate formulation must place either explicit or implicit constraints upon the causal structure of credal evolution. I cannot prove to you that the mediation clause is needed. But I

---

[9] Williamson's proof requires the assumption that, for all propositions H, $P_{new}(H) = 1$ if $P_E(H) = 1$. The assumption is satisfied here, since E is the strongest proposition F such that $P_{new}(F) = 1$ *and* the strongest proposition F such that $P_E(F) = 1$. Diaconis and Zabell (1982) prove a result along the same lines for probability measures over subsets of a countable outcome space. In this setting, the requisite partition $\{E_i\}$ may be countably infinite.

challenge you to provide an adequate formulation that avoids explicit or implicit appeal to mediation.

To illustrate the importance of causal structure, consider how Meacham (2016, p. 784) formulates Conditionalization:

> It should^{s,t_1} be the case that if a subject *s* with credences *cr* at t_0 gets cumulative evidence *E* in the [t_0; t_1] interval, then she will adopt new credences cr_E at t_1 such that $cr_E(\, . \,) = cr(\, . \mid E)$, if defined.

The indices *s* and t_1 on "should" register that deontic obligation may vary with the agent and the time. Meacham does not say what "evidence" is. As he acknowledges (pp. 767-768), his formulation is therefore schematic. It does not have a determinate fulfillment-condition. To specify a determinate fulfillment-condition, one must say what evidence is.

Can credal shifts over a partition count as evidence? If not, then Meacham's schematic formulation is vulnerable to a Jeffrey-style counterexample. An agent who gains cumulative evidence *E* may *also* undergo a Jeffrey-style shift. Such an agent should not conditionalize on *E*.

On the other hand, suppose that evidence *can* include credal shifts over a partition. Meacham (2016, p. 769) recommends that we allow this, partly in order to accommodate Jeffrey-style shifts. We now face a further question: *which* credal shifts count as evidence? Meacham does not say. Suppose he were to respond that *any* credal shift over *any* partition counts as evidence. The proposed response renders Conditionalization trivial. Given any credal change, we can always find a partition over which credences shift. For example, an agent who conditionalizes on *E* will shift her credences over the partition $\{E \,\&\, F, \neg(E \,\&\, F)\}$, where *F* is some proposition that is probabilistically independent of *E*. (We may take *F* to describe the result of a coin toss that is unrelated to *E*.) The proposed response entails that "cumulative evidence" is never exhausted by newfound certainty in a proposition *E*, so that Conditionalization's antecedent is never satisfied, so that Conditionalization is always trivially satisfied. Meacham must therefore grant that some credal shifts over a partition do *not* count as evidence.

Which credal shifts count as evidence? When is an agent's evidence exhausted by newfound certainty in a proposition, and when does the agent's evidence include new credal assignments to partition propositions? I see no satisfying way for Meacham to answer these questions without constraining the causal structure of the agent's credal evolution. A credal shift across the partition $\{E \,\&\, F, \neg(E \,\&\, F)\}$ should not count as evidence when it results from the agent's newfound certainty in *E*. The credal shift in §4's Jeffrey-style counterexample might potentially count as evidence, because it is the source rather than the consequence of all other credal changes. Thus, mere appeal to "evidence" is no panacea regarding Jeffrey-style counterexamples. One must still find some way to delimit those scenarios where credal evolution is sparked *solely* by newfound certainty in a proposition. A suitable constraint on causal structure seems needed. One can include the constraint explicitly in the statement of Conditionalization, or one can include it implicitly by building it into one's conception of "evidence." Either way, something like the mediation clause plays a crucial role in blocking the counterexamples.

§6.3 Oblique evocation of causal structure

Many formulations of Conditionalization obliquely evoke causal structure while eschewing explicitly causal language. For example, Joyce (2009, p. 419) says that an agent should conditionalize on *x* when "the only new information she acquires is that *x* is certainly true." He does not explain the key phrase "the only new information she acquires," so this formulation is difficult to interpret. Maybe he has in mind the same intuitive idea that I have emphasized: a credal reallocation event alters the agent's credences entirely *by way of* shifting all probability mass to a single member of some partition. Better to make this intuitive idea explicit through overtly causal vocabulary.

Explicitly causal vocabulary figures more commonly in formulations of Jeffrey Conditionalization. When stating Jeffrey Conditionalization, Earman (1992, p. 34) posits that credal changes are "generated" by new credences over the partition $\{E_i\}$. Similarly, Joyce imagines an experience whose "only immediate effect" is to fix new credences over the partition (2009, pp. 35-36).[10] I submit that we require kindred causal locutions when formulating ordinary Conditionalization. Indeed, to the extent that we regard ordinary Conditionalization as a special case of Jeffrey Conditionalization, causal discourse should figure in our formulation of the former precisely to the same extent that it figures in our formulation of the latter.[11]

Some authors seek to handle Jeffrey-style counterexamples by incorporating causal language into *background remarks* rather than the official formulation of Conditionalization. For example, Titelbaum formalizes Conditionalization as follows (2013, p. 124):

For any t_j, $t_k \in T$ with $j \leq k$ and any $x \in L$, if $C_k \subseteq C_j$ then $P_k(x) = P_j\ (x \mid \langle C_k - C_j \rangle)$,

where T is a set of times; L is a modeling language whose sentences serve as objects of credence within Titelbaum's framework; P_k and P_j are credal allocation functions for times k and j respectively; C_k is the set of all sentences that are certain at time k, and similarly for C_j; and $\langle C_k - C_j \rangle$ is a single sentence that captures everything said by all the sentences belonging to C_k but not

[10] Elsewhere, Joyce (2019) uses similar causal language to formulate Conditionalization. He says that you should conditionalize on E when you have "a learning experience whose sole immediate effect" is to render you certain of E. This formulation is relatively close to my own: it uses causal language, and it seems to be non-factive. However, it does not handle modified Jeffrey-style counterexamples along the lines of Figure 11.
[11] When explaining Jeffrey Conditionalization, Jeffrey (1983, p. 173) assumes that the change from P_{old} to P_{new} "originates" in the partition propositions $\{E_i\}$. He then defines "origination" in terms of the invariance condition: $P_{old}(. \mid E_i) = P_{new}(. \mid E_i)$, for all i. Jeffrey's discussion epitomizes how authors often recognize that causal structure is important yet balk at incorporating explicitly causal vocabulary into the official formulation of Conditionalization or Jeffrey Conditionalization.

to C_j.[12] As Titelbaum (2013, pp. 104-105) admits, his formalization does not begin to prevent misapplication to Jeffrey-style shifts. In surrounding remarks, he clarifies that his formalization is "designed to model degree of belief changes *driven* by changes in certainties" (2013, p. 105). He does not explicate "driven," but the intuitive idea seems close to my notion of mediation. Plainly, the credal changes in a Jeffrey-style shift are not all "driven" by changes in certainties. Ultimately, then, Titelbaum deploys something resembling the mediation clause, situated not within his official statement of Conditionalization but rather within background remarks that delimit Conditionalization's intended domain of application.

I think it preferable whenever possible to incorporate restrictions on a norm's domain of application into the norm itself, rather than appending them as ancillary background remarks. Conditionalization is a substantive diachronic norm, not just an equation in a formal calculus. We should specify its fulfillment-condition as perspicuously as possible. To do so, we should delimit its domain of application *within the content of the norm* as much as possible. Sure, we can achieve deceptive surface clarity by relocating causal locutions from our official formulation of Conditionalization into background remarks. But we gain this surface clarity only by shunting causal locutions from view, thereby disguising their pivotal role in delimiting Conditionalization's intended domain of application. No overall increase in clarity is achieved. Consigning causal locutions to background remarks also encourages the harmful misimpression that those locutions require no systematic elucidation. Whereas some authors obliquely evoke unexplicated causal aspects of Conditionalization, I explicitly highlight all causal aspects and elucidate them in relatively systematic terms.

[12] Titelbaum calls this requirement "Limited Conditionalization." He reserves the label "Conditionalization" *simpliciter* for a similar requirement that omits the subset clause $C_k \subseteq C_j$ (p. 52). He also articulates a generalized requirement, "Generalized Conditionalization," designed to cover the cases of memory loss and apparent memory loss discussed in §7 below. None of Titelbaum's three formulations attempts to handle Jeffrey-style shifts. Thus, my criticism in the main text applies equally well to all three formulations.

§7. Mediation and memory loss

An important test for any formulation of Conditionalization is how well it handles various problem cases. In that spirit, let us consider how my proposed formulation handles some widely discussed scenarios involving *memory loss*.

Talbott (1991) explores a scenario along the following lines. On July 15, 2015, you are certain that you ate granola for breakfast that day. A year later, you forget what you ate for breakfast on July 15, 2015. Your credence in the proposition *I ate granola for breakfast on July 15, 2015* has dropped far below 1. Conditionalizing on a proposition E with $P_{old}(E) > 0$ can never dislodge credences from 1 after they have settled there. An agent who updates her beliefs solely as dictated by Conditionalization cannot lose certainties. So a framework where Conditionalization is the *only* credal reallocation norm cannot model Talbott's memory loss scenario. Similarly for Jeffrey Conditionalization.

Some authors claim that memory loss scenarios *violate* Conditionalization. Eagle writes (2011, p. 123): "If an agent forgets something they used to know, they will clearly violate both regular conditionalization and Jeffrey conditionalization. For if what they knew had credence 1, and since no conditionalizing update will ever reduce the credence in any proposition which has credence 1, then agents who update only by conditionalization never forget." I reply that we must distinguish between scenarios that *violate* a norm and scenarios that *trivially satisfy* a norm because the norm's antecedent is false. Failure to conditionalize does not entail that you violate Conditionalization, so long as you are operating in circumstances where Conditionalization's antecedent is false. In Talbott's example, COND's antecedent is not satisfied by any partition $\boldsymbol{E}$, because the transition from P_{old} to P_{new} is not mediated by new certainties. At least some of your credences change through a quite different causal mechanism: memory loss. Thus, Talbott's memory loss scenario is not a counterexample to Conditionalization *as properly formulated*. When COND's antecedent is false, COND yields no instructions about how to reallocate credence. This is a problem for a view on which conditionalizing is the only rational method for updating credences. And it may be that certain Bayesians have held that extreme view. However, endorsing Conditionalization does not commit one to anything approaching that extreme view. For all COND says, there may be many further norms that govern credal reallocation in various scenarios. Many operations besides conditionalizing may be rational methods for updating credences in suitable circumstances.[13]

Arntzenius (2003) advances another widely discussed example involving memory loss: *the road to Shangri La*. A coin toss will decide how you travel to Shangri La: heads, you go by the mountains; tails, you go by the sea. If you go by the mountains, you will remember that you went by the mountains. If you go by the sea, the guardians of Shangri La will change your memories upon arrival so that you *seem* to remember going by the mountains. Either way, your apparent memories will be exactly the same once you arrive in Shangri La. You know all this with complete certainty. The coin lands heads, so you go by the mountains. What credence should you invest in the proposition *The coin landed heads* throughout your peregrinations?

Arntzenius contends that your credence should be 1 as you are traveling by the mountains and ½ once you arrive in Shangri La. Let us grant that Arntzenius is correct. Since conditionalizing on a proposition E with $P_{old}(E) > 0$ cannot eliminate certainties, it follows that your credences once settled in Shangri La do not arise from conditionalizing on any such E.

[13] Here we see a major advantage of my formulation over Meacham's evidence-based formulation. Whatever conception of "evidence" one adopts, memory loss cases look like counterexamples to Meacham's formulation. Take Talbott's breakfast example, and let E be the cumulative evidence acquired between July 15, 2015 and July 15, 2016. Meacham's formulation requires that your new credences on July 15, 2016 be given by conditionalizing on E. Conditionalizing on E would yield credence 1 in the proposition *I ate granola for breakfast on July 15, 2015*, which seems inappropriate given that you cannot remember what you ate for breakfast that day. Thus, Meacham's approach yields implausible results in memory loss cases.

Arntzenius (p. 356) concludes that rational agents can "violate Bayesian conditionalization." He explains: "Standard Bayesian lore has it that rational people satisfy the principle of conditionalization: rational people alter their degrees of belief only by strict conditionalization on the evidence acquire" (pp. 366-367). He does not cite any authors who endorse the enunciated "principle of conditionalization." In any event, the Shangri La example is not a counterexample to Conditionalization *as properly formulated*. The transition from P_{old} (your credal allocation while traveling to Shangri La) to P_{new} (your credal allocation once settled in Shangri La) is not mediated by acquisition of new certainties. It is triggered partly by *loss* of certainty in the proposition *The coin landed heads* (maybe along with loss of certainty in additional propositions, such as a proposition attesting that your apparent memories are veridical). The antecedent of COND is not satisfied by any partition $\boldsymbol{E}$, so COND issues no instructions regarding how to reallocate credence. Rational agents in the Shangri La scenario do not violate Conditionalization as formulated through COND.

The Shangri La example is intriguing partly because it suggests that certainties can be rationally eradicated not just through memory loss but also through the mere threat of memory loss. Still, the scenario is no more a counterexample to Conditionalization *as properly formulated* than is Talbott's simpler memory loss scenario.

The question remains: can we model memory loss and threatened memory loss within a broadly Bayesian framework? Clearly, doing so would require further update rules beyond Conditionalization and Jeffrey Conditionalization. This is an important topic, but it falls outside the scope of my discussion. Meacham (2008) and Titelbaum (2013) explore how one might generalize the Bayesian framework to handle such cases.

§8. Conditionalization as a norm of credal evolution

I conclude by highlighting two key features of my preferred formulation COND. These features distinguish my approach from many, if not most, formulations found in the literature.

§8.1 Wide scope

Conditionalization is a norm that constrains credal evolution. We may schematize the norm as follows:

(16) □(If *p*, then *q*),

where the box is a deontic operator meaning "it is required that" and the conditional inside the box schematizes COND. Here I suppress quantificational structure that any full schematization would reveal. The deontic operator in (16) has *wide scope*: it governs the entire conditional. We must distinguish (16) from

(17) If *p*, then □*q*,

where the deontic operator has *narrow scope*. (17) says that, if certain conditions are fulfilled, then P_{new} should have certain properties. In contrast, my preferred formulation of Conditionalization does not say that P_{new} should have certain properties. It does *not* say:

> If you begin with credences P_{old}, and $\boldsymbol{E} = \{E_i\}$ is a countable set of mutually exclusive, jointly exhaustive propositions such that $P_{old}(E_i) > 0$ for each i, and you subsequently adopt new credences P_{new} such that $P_{new}(E_k) = 1$ and $P_{new}(E_i) = 0$ for each $i \neq k$, and the new credal assignment over $\boldsymbol{E}$ mediates the transition from P_{old} to P_{new}, then it is required that $P_{new}(\,.\,) = P_{old}(\,.\mid E)$.

See (Titelbaum, 2013, pp. 67-70) for explication of the distinction between wide and narrow scope formulations of Conditionalization.

Authors often formulate Conditionalization as a narrow scope norm rather than a wide scope norm (e.g. Hájek, 2009, p. 165; Joyce, 2009, p. 419; Ramsey, 1931/1990, p. 88; Weisberg, 2009, p. 499; Talbott, 2015). Consider Huber's (2016) formulation:

> If evidence comes only in the form of *certainties* (that is, propositions of which you become certain), if Pr: $\mathbf{A} \rightarrow \mathbb{R}$ is your subjective probability at time t, and if between t and t' you become certain of $A \in \mathbf{A}$ and no logically stronger proposition in the sense that your new subjective probability for A, but for no logically stronger proposition, is 1 (and your subjective probabilities are not directly affected in any other way such as forgetting etc.), then your subjective probability at time t' should be $\Pr(\cdot \mid A)$,

where $\mathbf{A}$ is the set of relevant propositions and $\mathbb{R}$ is the set of real numbers. In many respects, Huber's non-factive formulation resembles mine. He is one of the few authors who explicitly includes a causal condition akin to the mediation clause ("your subjective probabilities are not directly affected in any other way"). However, Huber's formulation differs from mine by assigning the deontic operator narrow scope.

I agree with Meacham (2016) and Titelbaum (2013, pp. 67-70) that a wide scope formulation of Conditionalization is more plausible than a narrow scope formulation. As Christensen (2000) observes, your initial credence P_{old} may be irrational. For example, you may assign high initial credence to an implausible conspiracy theory for which you have no good evidence. When P_{old} is irrational, Christensen denies that you should adopt new credences derived from P_{old} by conditionalizing on a proposition E. Wide scope formulations of Conditionalization avoid this worry (Titelbaum 2013, p. 69), because they do not say that you should conditionalize on E in this or any other circumstance. Similarly, suppose you begin with reasonable credences P_{old} but then newly assign credence 1 to an implausible proposition E

based on no good evidence, e.g. due to a bump on the head. It is hardly clear that you should conditionalize on E. Wide scope formulations avoid this worry, because they do not say that you should conditionalize on E.

A wide scope formulation brings into the focus the true nature of Conditionalization. As Titelbaum (2013, pp. 67-70) emphasizes, Conditionalization is most fundamentally a norm governing credal *evolution*. The basic unit of assessment is not P_{new} itself, or P_{old} and P_{new} taken together, but rather the *psychological process* through which the subject's credences evolve from P_{old} to P_{new}. Conditionalization requires that the process have certain features. A process that lacks those features is rationally defective. If the credal evolution from P_{old} to P_{new} is rationally defective, it does not follow that P_{new} itself is defective. What follows is just that *something* has gone wrong, rationally speaking, in the process that begins with P_{old} and ends with P_{new}.

§8.2 Non-credal epistemic states?

When formulating Conditionalization, authors often characterize the relation between agents and conditioning propositions using informal epistemic locutions, such as "knowledge," "evidence," "learning," "information," "observation," and "experience." These locutions are *non-credal*, in that they do not explicitly mention credal assignments. By using non-credal descriptions, authors convey the impression that Conditionalization concerns how the agent's credal allocations interface with certain non-credal epistemic states.

My proposed formulation eschews non-credal language when describing how agents relate to conditioning propositions. It is far from clear how COND bears upon any norm framed in terms of "knowledge," "evidence," "learning," "information," "observation," "experience," or other non-credal epistemic locutions. On my approach, Conditionalization does not directly address how credal assignments interface with non-credal epistemic states. It does not directly address how agents should reallocate credence upon acquiring new knowledge, learning a new proposition, or gaining new evidence. Most fundamentally, Conditionalization constrains credal evolution sparked by an exogenous shift in probability mass to a single proposition. Conditionalization places an *internal* constraint on the dynamics of credal evolution, not an *external* constraint on the interface between credal evolution and non-credal states.

One advantage of my approach is that it covers certain situations that we otherwise risk overlooking. For example, suppose you take a pill that impacts your credences by causing you to set $P_{new}(E) = 1$ and $P_{new}(\neg E) = 0$. In this situation, you do not come to know *E*. You do not learn *E*. You do not gain evidence that *E*. Nevertheless, we want a norm that governs your credal evolution. COND does so, unlike most rival formulations found in the literature.

Some authors who use informal epistemic locutions to formulate Conditionalization may intend those locutions as heuristic proxies for talk about credal allocation. Some authors may envisage a reduction of non-credal epistemic states to credal states. Of course, any such reduction would be controversial. An optimal formulation of Conditionalization should eschew heuristic paraphrases and controversial theses whenever possible. Non-credal epistemic locutions are perhaps pedagogically useful when introducing Conditionalization to novices, but they have no place in a rigorous statement of what Conditionalization requires. An exception is the locution

X is certain that *p*,

which many Bayesians use stipulatively to mean

X assigns credence 1 to *p*.

This stipulative usage arguably captures one important notion of "certainty." Setting aside certainty, I insist that informal epistemic locutions should not figure in a proper formulation of Conditionalization.

Contemporary philosophers often apply the Bayesian framework to traditional epistemological questions. It may be that Conditionalization has important implications for traditional informal epistemology. Most fundamentally, though, Conditionalization does not concern knowledge, warrant, justification, or evidence. It concerns the proper evolution of credences in response to new certainties (possibly misplaced certainties, possibly certainties with no evidentiary basis). We should resist unhelpful incursions of informal epistemology into the foundations of Bayesian decision theory.

§9. A plea for pedantry

Philosophers are supposed to be pedantically obsessed with formulating claims as carefully and precisely as possible. In the case of Conditionalization, there has not been enough pedantry. Existing formulations are often quite sloppy. Many trade ambiguously between factive and non-factive readings. Many employ unexplicated, unhelpful epistemic locutions. Some evoke causal aspects of Conditionalization using covertly causal locutions (e.g. "driven," "generated"), often consigned to ancillary background remarks. There is an all-too-common tendency to proceed as if equation (2) is itself a proper statement of Conditionalization, with any additional commentary only so much surrounding heuristic. As I have stressed, Conditionalization is a *norm* that has a reasonably determinate fulfillment-condition. We should strive to articulate this fulfillment-condition as carefully as possible.

No doubt my proposed formulation could withstand improvement. I urge you to propose your own improved formulation. But I implore you when doing so to enforce normal philosophical standards of precision and rigor --- standards that lapse all too frequently during discussion of Conditionalization.

Acknowledgments

I presented versions of this material at Carnegie Mellon University and Stanford University. I thank all the audience members present on those occasions, especially Dmitri Gallow, Clark Glymour, Thomas Icard, and Teddy Seidenfeld. I also thank Christopher Meacham and two anonymous referees for this journal for comments that improved the paper significantly.

Works Cited

Arntzenius, F. 2003. “Some Problems for Conditionalization and Reflection.” *The Journal of Philosophy* 100: pp. 356-370.
Berger, J. 1980. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.
Briggs, R. 2009. “Distorted Reflection.” *Philosophical Review* 118: 59-85.
Christensen, D. 2000. “Diachronic Coherence versus Epistemic Impartiality.” *Philosophical Review* 109: pp. 349-371.
de Finetti, B. 1970/1974. *Theory of Probability*, vol. 1, trans. A. Machi and A Smith. New York: Wiley.
Diaconis, P., and Zabell, S. 1982. “Updating Subjective Probability.” *The Journal of the American Statistical Association* 77: pp. 822-830.
Eagle, A. 2011. “Introduction.” In *Philosophy of Probability: Contemporary Readings*, ed. A. Eagle. New York: Routledge.
Earman, J. 1992. *Bayes or Bust?* Cambridge: MIT Press.
Easwaran, K. 2011. “Bayesianism I: Introduction and Arguments in Favor.” *Philosophy Compass* 6: 312-320.
Fudenberg, D., and Tirole, J. 1991. *Game Theory*. Cambridge: MIT Press.
Greaves, H., and Wallace, D. 2006. “Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility.” *Mind* 115: pp. 607-632.

Hacking, I. 1967. “Slightly More Realistic Personal Probability.” *Philosophy of Science* 34: pp. 311-325.

Hall, N. 2004. “Two Concepts of Causation.” In *Causation and Counterfactuals*, eds. J. Collins, N. Hall, and L. A. Paul. Cambridge: MIT Press.

Halpern, J., and Hitchcock, C. 2015. “Graded Causation and Defaults.” *The British Journal for the Philosophy of Science* 66: pp. 413-457.

Hájek, A. 2009. “Dutch Book Arguments.” In *The Handbook of Rational and Social Choice*, eds. P. Anand, P. Pattanaik, and C. Puppe. Oxford: Oxford University Press.

Howson, C., and Franklin, A. 1994. “Bayesian Conditionalization and Probability Kinematics.” *The British Journal for the Philosophy of Science* 45: pp. 451-466.

Huber, F. 2016. “Formal Representations of Belief.” *The Stanford Encyclopedia of Philosophy* (Spring 2016), ed. E. Zalta. http://plato.stanford.edu/archives/spr2016/entries/formal-belief/.

Jeffrey, R. 1983. *The Logic of Decision*. 2nd edition. Chicago: University of Chicago Press.

---. 2004. *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press.

Jeffreys, H. 1948. *Theory of Probability*, 2nd ed. Oxford: Clarendon Press.

Joyce, J. 2009. “The Development of Subjective Bayesianism.” In *Handbook of the History of Logic*, vol. 10, eds. D. Gabbay, S. Hartman, and J. Woods. New York: Elsevier.

---. 2019. “Bayes’ Theorem.” *The Stanford Encyclopedia of Philosophy* (Spring 2019). ed. E. Zalta. http://plato.stanford.edu/archives/win2016/entries/bayes-theorem/.

Lewis, D. 1986. “Causation.” In *Philosophical Papers*, vol. 2. Oxford: Oxford University Press.

---. 1999. *Papers in Metaphysics and Epistemology*. Cambridge: Cambridge University Press.

Meacham, C. 2008. “Sleeping Beauty and the Dynamics of *De Se* Beliefs.” *Philosophical Studies* 138: pp. 245-270.

---. 2016. “Understanding Conditionalization.” *Canadian Journal of Philosophy* 45: pp. 767-797.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Ramsey, F. P. 1931/1990. “Truth and Probability.” Rpt. in *Philosophical Papers*, ed. D. H. Mellor. Cambridge: Cambridge University Press.

Rescorla, M. 2014. “The Causal Relevance of Content to Computation.” *Philosophy and Phenomenological Research* 88: pp. 173-208.

---. Forthcoming. “A Dutch Book Theorem and Converse Dutch Book Theorem for Kolmogorov Conditionalization.” *The Review of Symbolic Logic*.

Savage, L. 1954/1972. *The Foundations of Statistics*. Dover: New York.

Schaffer, J. 2005. “Contrastive Causation.” *Philosophical Review* 114: pp. 297-328.

Schoenfield, M. 2017. “Conditionalization Does Not (in general) Maximize Expected Accuracy.” *Mind* 504: pp. 1155-1187.

Skyrms, B. 1987. “Dynamic Coherence and Probability Kinematics.” *Philosophy of Science* 54: 1–20.

Talbott, W. 1991. “Two Principles of Bayesian Epistemology.” *Philosophical Studies* 62: pp. 135-150.

---. 2015. “Bayesian Epistemology.” *The Stanford Encyclopedia of Philosophy* (Summer 2015), ed. E. Zalta. http://plato.stanford.edu/archives/sum2015/entries/epistemology-bayesian/.

Titelbaum, M. 2013. *Quitting Certainties*. Oxford: Oxford University Press.

van Fraassen, B. 1999. “Conditionalization, A New Argument For.” *Topoi* 18: pp. 93-96.

Vineberg, S. 2011. “Dutch Book Arguments.” *The Stanford Encyclopedia of Philosophy* (Summer 2011), ed. E. Zalta. http://plato.stanford.edu/archives/sum2011/entries/dutch-book/.

Weisberg, J. 2009. “Varieties of Bayesianism.” In *Handbook of the History of Logic*, vol. 10, eds. D. Gabbay, S. Hartman, and J. Woods. New York: Elsevier.

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

Woodward, J. 2003. *Making Things Happen*. Oxford: Oxford University Press.

---. 2015. “Interventionism and Causal Exclusion.” *Philosophy and Phenomenological Research* 91: pp. 303-347.

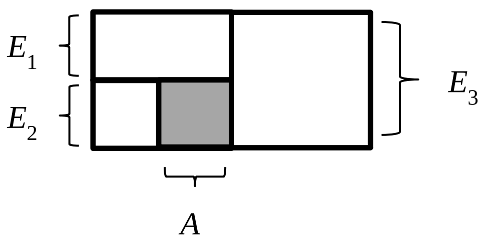

Figure 1. Propositions E_1, E_2, and E_3 partition the probability space. The size of each rectangle is proportional to the probability mass initially assigned to the corresponding proposition. Since $P(A \mid E_2) = ½$, the A rectangle occupies half of the E_2 rectangle.

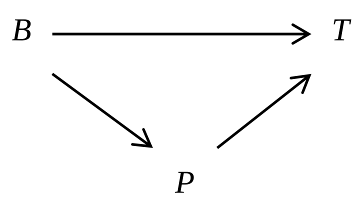

Figure 2. *B* is directly causally relevant to *T* relative to $V = \{B, P, T\}$. There is also an indirect causal route from *B* to *T* via *P*.

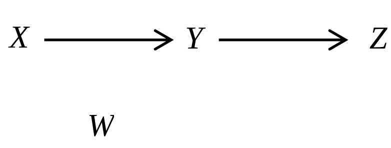

Figure 3. *X* is causally relevant to *Y*. *Y* is causally relevant to *Z*. *X* is not directly causally relevant to *Z* relative to $V = \{W, X, Y, Z\}$. *W* is not causally relevant to *X*, *Y*, or *Z*.

Figure 4. A simple causal structure.

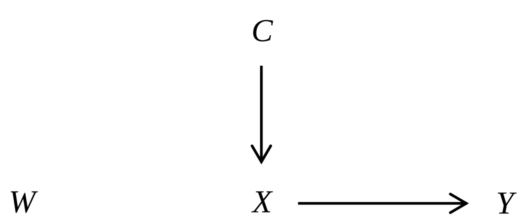

Figure 5. Compare with Figure 4. An intervention “breaks” the arrow from *W* to *X*, replacing the causal structure from Figure 4 with a new causal structure.

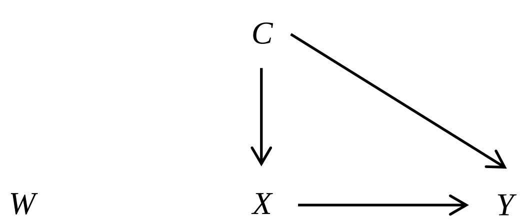

Figure 6. This causal structure violates clause I.4

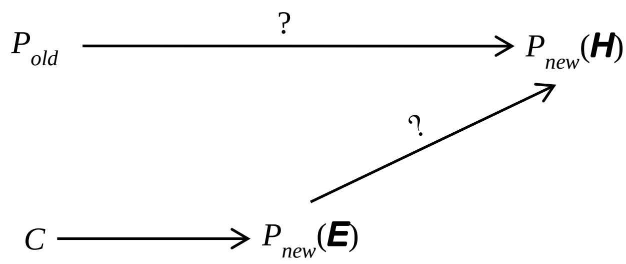

Figure 7. A causal structure that conforms to (10). C controls $P_{new}(\boldsymbol{E})$, so there is no arrow from P_{old} to $P_{new}(\boldsymbol{E})$. C is not directly causally relevant to $P_{new}(\boldsymbol{H})$ relative to $\{C, P_{new}(\boldsymbol{E}), P_{new}(\boldsymbol{H})\}$, so there is no arrow from C to $P_{new}(\boldsymbol{H})$. The question marks serve to leave open whether the corresponding arrows are present, e.g. whether P_{old} is causally relevant to $P_{new}(\boldsymbol{H})$. (10) allows but does not require that the corresponding arrows are present. (10) also allows that there may be an arrow from C to P_{old}. I have omitted any arrow from C to P_{old} because, in all cases of interest to us, there exists an exogenous variable C that influences $P_{new}(\boldsymbol{E})$ but not P_{old}.

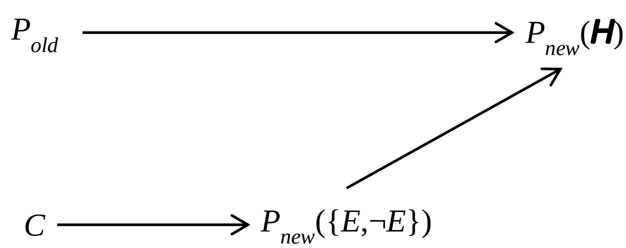

Figure 8. A causal structure instantiated when an agent conditionalizes on E in response to newfound certainty in E.

$$\begin{tikzcd} P_{old} \arrow[rr] & & P_{new}(\boldsymbol{H}) \\ C \arrow[r] & P_{new}(\{E_i\}) \arrow[ur] & \end{tikzcd}$$

Figure 9. A causal structure instantiated when an agent Jeffrey conditionalizes in response to new credences over $\{E_i\}$.

$$\begin{tikzcd} P_{old} \arrow[rr] & & P_{new}(E_1 \vee E_2) \\ C \arrow[r] & P_{new}(\{E_1, E_2, E_3\}) \arrow[ur] & \end{tikzcd}$$

Figure 10. A causal structure instantiated by the Jeffrey-style counterexample from §4.

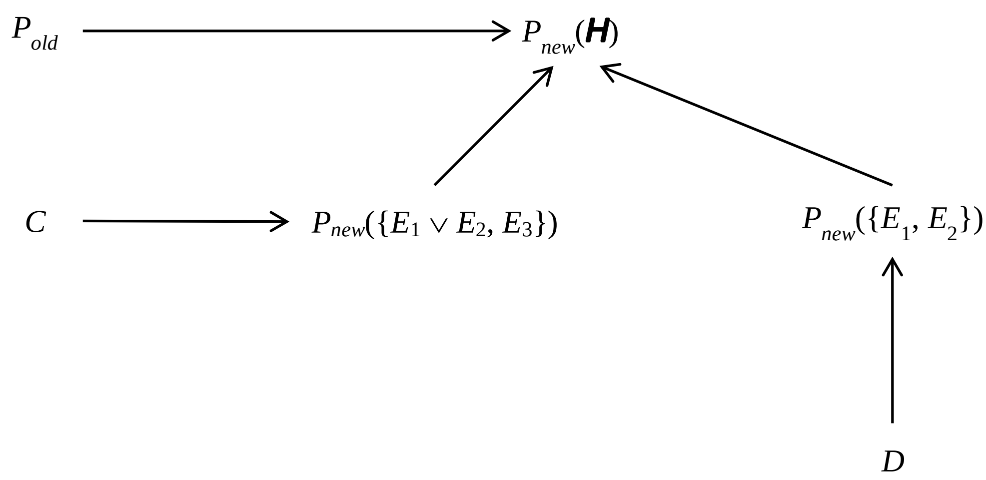

Figure 11. A causal structure instantiated by our modified Jeffrey-style counterexample.

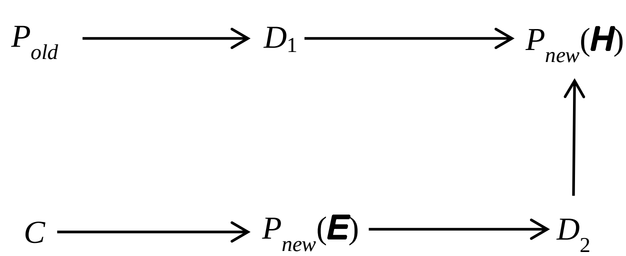

Figure 12. A causal structure featuring two "intermediate variables" D_1 and D_2.

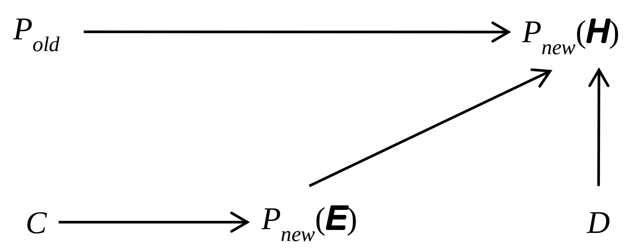

Figure 13. One possible way that *D* might satisfy clauses M.2.c-M.2.e.

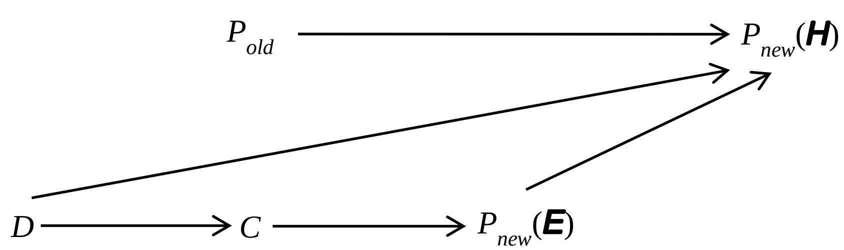

Figure 14. Another possible way that *D* might satisfy clauses M.2.c-M.2.e.