



Bayesian defeat of certainties

Michael Rescorla¹ 

Received: 7 April 2023 / Accepted: 11 October 2023 / Published online: 31 January 2024
© The Author(s) 2024

Abstract

When $P(E) > 0$, conditional probabilities $P(H|E)$ are given by the ratio formula. An agent engages in *ratio conditionalization* when she updates her credences using conditional probabilities dictated by the ratio formula. Ratio conditionalization cannot eradicate certainties, including certainties gained through prior exercises of ratio conditionalization. An agent who updates her credences only through ratio conditionalization risks permanent certainty in propositions against which she has overwhelming evidence. To avoid this undesirable consequence, I argue that we should supplement ratio conditionalization with *Kolmogorov conditionalization*, a strategy for updating credences based on propositions E such that $P(E) = 0$. Kolmogorov conditionalization can eradicate certainties, including certainties gained through prior exercises of conditionalization. Adducing general theorems and detailed examples, I show that Kolmogorov conditionalization helps us model epistemic defeat across a wide range of circumstances.

Keywords Bayesian decision theory · Conditional probability · Regular conditional distribution · Conditionalization · Epistemic defeat

1 Beyond ratio conditionalization

Bayesian decision theory studies an idealized rational agent who assigns *subjective probabilities*, or *credences*, to propositions. The agent's credences at each moment conform to the probability calculus axioms. The agent *conditionalizes on E* when she replaces her initial credences $P_{old}(H)$ with new credences $P_{new}(H)$ given by

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

✉ Michael Rescorla
rescorla@ucla.edu

¹ Department of Philosophy, University of California, Los Angeles, USA

where $P_{old}(H|E)$ is the initial conditional probability of H given E . If $P_{old}(E) > 0$, then $P_{old}(H|E)$ is given by the familiar *ratio formula*:

$$P_{old}(H|E) = \frac{P_{old}(H \& E)}{P_{old}(E)},$$

yielding the following formula for P_{new} :

$$P_{new}(H) = \frac{P_{old}(H \& E)}{P_{old}(E)}. \quad (1)$$

When an agent updates her probabilities in accord with (1), I will say that she engages in *ratio conditionalization*.

Suppose that an agent is *certain of H* , in the sense that she sets $P_{old}(H) = 1$. If $P_{old}(E) > 0$, then the ratio formula and the probability calculus axioms entail that $P_{old}(H|E) = 1$. It follows that ratio conditionalization cannot eliminate certainties. An agent who is certain of H and who updates her credences solely through ratio conditionalization will remain forever certain of H .

A striking illustration arises through the very exercise of ratio conditionalization. The ratio formula and the probability calculus axioms entail that $P_{old}(E|E) = 1$, so an agent who conditionalizes on E must set $P_{new}(E) = 1$. She must also set $P_{new}(E|F) = 1$ for any F such that $P_{new}(F) > 0$. Thus, subsequent exercises of ratio conditionalization cannot dislodge her newfound certainty in E , no matter what further evidence F she receives. This situation is widely regarded as disturbing (Jeffrey, 1983; Levi, 1980; Titelbaum, 2013; Williamson, 2000; Weisberg, 2009b). Surely an agent who conditionalizes on E can later receive strong evidence against E ! For example, a scientist may conditionalize on the proposition that her experiment had a certain outcome and later learn that the experiment had a different outcome (e.g. her laboratory assistant initially misreported an experimental measurement). An agent who updates her credences only through ratio conditionalization risks permanent certainty in propositions against which she has overwhelming evidence.

Partly in response to such worries, Jeffrey (1983) develops an alternative framework that de-emphasizes conditionalization. Jeffrey contends that empirical propositions rarely if ever merit credence 1 and hence that one should rarely if ever update using ratio conditionalization. He proposes an update strategy, now usually called *Jeffrey conditionalization*, that does not mandate new certainties. Jeffrey conditionalization applies when an external influence causes an agent to reallocate credences across a partition containing countably many mutually exclusive, jointly exhaustive propositions $E_1, E_2, \dots, E_i, \dots$. Jeffrey proposes that, in response to the reallocated credences, the agent should form new credences given by:

$$P_{new}(H) = \sum_i P_{old}(H|E_i)P_{new}(E_i). \quad (2)$$

An agent who uses Jeffrey conditionalization rather than ratio conditionalization can update her credences without acquiring new certainties.

I think that Jeffrey conditionalization is a useful update strategy in some circumstances. However, I doubt that it can completely replace ordinary conditionalization. Ordinary conditionalization figures crucially within scientific applications of the Bayesian framework. Wholesale replacement of ordinary conditionalization by Jeffrey conditionalization would require sweeping changes to scientific practice, with major side effects both foreseen and unforeseen. It is far from clear that those changes would preserve the explanatory and pragmatic achievements of contemporary Bayesian practice (Rescorla, 2022).

In any event, I will follow a different path. The basic idea I will pursue is that an agent can eradicate certainties by conditionalizing on an E such that $P_{old}(E) = 0$. Although $P_{old}(H|E) = 1$ whenever $P_{old}(H) = 1$ and $P_{old}(E) > 0$, there is no reason to expect that $P_{old}(H|E) = 1$ when $P_{old}(H) = 1$ and $P_{old}(E) = 0$. On the contrary, one would expect that $P_{old}(H|E) < 1$ for any H that entails $\neg E$. In principle, then, an agent who conditionalizes on probability zero propositions can lose certainties—including certainties gained through previous exercises of conditionalization.

Any account along these lines must look beyond the ratio formula to secure the needed conditional probabilities. Among the options found in the literature, I favor an approach that traces back to Kolmogorov (1933/1956) and that centers upon the notion of a *regular conditional distribution* (rcd). Rcds are central to probability theory. They also underlie countless scientific applications, especially within Bayesian statistics. Only recently have their virtues begun to receive sustained philosophical attention (Easwaran, 2008, 2011, 2019; Gyenis & Rédei, 2017; Huttegger, 2015; Huttegger & Nielsen, 2020; Meehan & Zhang, 2020, 2022; Nielsen, 2021; Rescorla, 2015b, 2018a, forthcoming). I will use them here to model certainty eradication across a range of situations.¹

Section 2 reviews basic aspects of rcds. Section 3 discusses how rcds support a kind of conditionalization, which I call *Kolmogorov conditionalization*. Sections 4 and 5 explore how Kolmogorov conditionalization can eradicate certainties. Section 6 articulates a rational norm, *Rigidity*, that constrains Kolmogorov conditionalization in many cases of epistemic defeat. Section 7 articulates a more general norm, *Generalized Rigidity*, that accommodates additional cases. The analysis from Sects. 6 and 7 indicates that rcds coupled with suitable rational norms can model numerous defeasible inferences. Section 8 highlights cases of epistemic defeat that my approach does *not* accommodate. Section 9 compares my approach with treatments due to Skyrms, Titelbaum, and Williamson.

2 Regular conditional distributions

Consider an idealized agent with prior credences modeled by a probability space (Ω, \mathcal{F}, P) , where Ω is a set, \mathcal{F} is a σ -field over Ω , and P is a probability measure on \mathcal{F} .² Elements of Ω are *outcomes*. Elements of \mathcal{F} , called *events*, serve as mathematical

¹ Several previous authors have suggested that one might model epistemic defeat of certainties by allowing conditionalization on an E such that $P_{old}(E) = 0$ (Pryor, 2013; Weisberg, 2009b). However, no one appears to have developed the suggestion in any detail, and no one has yet drawn the connection with rcds.

² See (Billingsley, 1995) for an introduction to measure-theoretic probability.

proxies for propositions. For each $A \in \mathcal{F}$, we construe $P(A)$ as the credence that the agent attaches to A . In a measure-theoretic setting, the ratio formula becomes:

$$P(H|E) = \frac{P(H \cap E)}{P(E)},$$

with intersection of events serving as a proxy for conjunction of propositions.

How should we define $P(H|E)$ when $P(E) = 0$? There is a pressing need for such conditional probabilities, quite independent of the considerations raised in Sect. 1. Scientific applications of Bayesian decision theory frequently require an agent to update her credences based upon learning that random variable X has value x . For example, an astronomer might learn that an asteroid has a certain velocity and, on that basis, update her credences regarding the time that the asteroid will reach Earth. Let

$$X = x$$

be shorthand for the event $\{\omega: X(\omega) = x\}$. If X has uncountably many possible values, then orthodox probability theory requires that $P(X = x) = 0$ for all but countably many values x (Billingsley, 1995, p. 188). Thus, the ratio formula cannot supply all the conditional probabilities $P(H|x = x)$ that we need.

Kolmogorov (1933/1956) offers a theory of conditional probability that goes far beyond the ratio formula. His central insight is that, when $P(E) = 0$, we should consider E not on its own but rather as embedded in a larger collection of events (some of which may also have probability zero). Formally speaking, Kolmogorov's theory centers on a subset $\mathcal{G} \subseteq \mathcal{F}$, where \mathcal{G} is itself a σ -field. His theory addresses scenarios where the agent gains new certainties over \mathcal{G} and on that basis reallocates credences over the rest of \mathcal{F} . New certainties over \mathcal{G} can be acquired through perception, testimony, introspection, or any other means. Some of the new certainties may only be implicit in the agent's mental activity. Kolmogorov constructs a systematic framework that models credal updates in light of implicit new certainties over \mathcal{G} .³

To understand Kolmogorov's approach, it helps to formalize the intuitive notion *implicit new certainties over \mathcal{G}* . For each $\omega \in \Omega$, define $\delta_\omega: \mathcal{G} \rightarrow \mathbb{R}$ by

$$\delta_\omega(G) = \begin{cases} 1 & \text{if } \omega \in G \\ 0 & \text{if } \omega \notin G \end{cases} \quad \text{for each } G \in \mathcal{G}.$$

Call δ_ω a *certainty profile* over \mathcal{G} , and call ω an *index* of δ_ω . Each certainty profile δ_ω models a scenario where the agent becomes certain that the true outcome does or does not belong to any given $G \in \mathcal{G}$.⁴ To illustrate, suppose the agent becomes certain that random variable X has value x . Assuming that she represents x through any standard notational scheme for the real numbers, she should be willing to affirm or deny that

³ Kolmogorov himself subscribed to a frequentist rather than subjectivist interpretation of probability. Thus, I do not claim that he *intended* to model credal updates in light of new certainties. I claim only that his mathematical framework serves this purpose admirably.

⁴ A single certainty profile can have many different indices. If outcomes ω and ν belong to precisely the same members of \mathcal{G} , then they index the same certainty profile: $\delta_\omega = \delta_\nu$.

X 's value lies between a and b , for each $a, b \in \mathbb{Q}$. Thus, she should be willing to assign probability 1 or 0 to each event

$$X^{-1}(a, b) \quad \text{for any } a, b \in \mathbb{Q}.$$

Let $\sigma(X)$ be the σ -field generated by these events, i.e. $\sigma(X)$ results from starting with the events $X^{-1}(a, b)$ and closing under complementation and countable union. The agent's new certainties over the events $X^{-1}(a, b)$ determine a unique certainty profile over $\sigma(X)$: namely, the certainty profile δ_ω , where ω is any outcome such $X(\omega) = x$. This certainty profile is implicit in the agent's newfound certainty that X has value x .

A certainty profile models a situation where the agent gains new certainties. In some cases, the agent's certainty profile δ_ω tracks the truth:

$$\text{For each } G \in \mathcal{G}, \delta_\omega(G) = 1 \text{ iff the true outcome belongs to } G. \tag{3}$$

I will call (3) *the factivity assumption*. Kolmogorov's framework accommodates scenarios where the factivity assumption prevails, and it also accommodates scenarios where the factivity assumption fails. In what follows, I will discuss scenarios of both kinds. As argued in (Rescorla, 2021), there is no principled reason to restrict attention to scenarios where the agent conditionalizes on truths. People make mistakes all the time. An agent's certainties over \mathcal{G} may be *misplaced*, i.e. they may violate the factivity assumption. Undoubtedly, though, it is a good thing when the factivity assumption prevails.

How should our agent reallocate credence over the rest of \mathcal{F} in light of her newfound certainty profile δ_ω ? To address this question, we will use a function $C: \mathcal{F} \times \Omega \rightarrow \mathbb{R}$. Intuitively, $C(\cdot, \omega)$ encodes probabilities over \mathcal{F} conditional on the truth of all those $G \in \mathcal{G}$ such that $\delta_\omega(G) = 1$. I will notate $C(A, \omega)$ as $C(A|\omega)$. When convenient, I will notate $C(\cdot|\omega)$ as C_ω . Our question now becomes what constraints we should place upon C . Since we are using probability measures to model credences, we demand that:

$$C_\omega : \mathcal{F} \rightarrow \mathbb{R} \text{ is a probability measure, for each } \omega \in \Omega. \tag{4}$$

Kolmogorov additionally demands that, for each $A \in \mathcal{F}$, the one-place function $C(A|\cdot): \Omega \rightarrow \mathbb{R}$ is \mathcal{G} -measurable:

$$C(A|\cdot)^{-1}(-\infty, a) \in \mathcal{G}, \text{ for each } a \in \mathbb{R}. \tag{5}$$

As explained in (Rescorla, forthcoming), \mathcal{G} -measurability formalizes the following intuitive thought: the agent's newfound certainties over \mathcal{G} dictate the new credences to be allocated over \mathcal{F} . Call any function C that satisfies clauses (4) and (5) an *update rule* for (Ω, \mathcal{F}) and \mathcal{G} .⁵

⁵ Assuming that C satisfies the \mathcal{G} -measurability requirement (5), C induces a well-defined mapping from certainty profiles to credences: if $\delta_\omega = \delta_\nu$, then $C(\cdot|\omega) = C(\cdot|\nu)$. See (Rescorla, forthcoming) for the proof, which is straightforward.

Kolmogorov supplements (4) and (5) with an additional constraint upon C . The constraint is now usually called the *integral formula*:

$$P(A \cap G) = \int_G C(A|\omega)dP(\omega), \text{ for each } A \in \mathcal{F} \text{ and } G \in \mathcal{G}. \tag{6}$$

The integral formula generalizes the *law of total probability*: for any partition $E_1, E_2, \dots, E_i, \dots$ such that $P(E_i) > 0$ for all i ,

$$P(A) = \sum_i P(A|E_i)P(E_i). \tag{7}$$

(7) follows from the ratio formula and the probability calculus axioms. Kolmogorov’s approach turns the entailment around, treating the law of total probability (generalized to the integral formula) not as a theorem but as a definitional constraint upon conditional probabilities.

A function $C: \mathcal{F} \times \Omega \rightarrow \mathbb{R}$ satisfying clauses (4)–(6) is called a *regular conditional distribution (rcd) for P given \mathcal{G}* . One can show that there exists an rcd for P given \mathcal{G} in a wide variety of cases, including virtually all cases likely to arise in scientific applications.⁶ I will frequently notate an rcd for P given \mathcal{G} as $P_{\mathcal{G}}$.

As a special case, suppose that \mathcal{G} is generated by a countable partition E_1, \dots, E_i, \dots , where $P(E_i) > 0$ for each i . Then there exists a unique rcd for P given \mathcal{G} , defined by:

$$P_{\mathcal{G}}(A|\omega) =_{df} P(A|E_i) = \frac{P(A \cap E_i)}{P(E_i)} \quad \text{if } \omega \in E_i.$$

In this way, Kolmogorov’s theory subsumes the ratio formula. Kolmogorov’s theory also supplies conditional probabilities in numerous cases where the ratio formula goes silent.

Rcds are widely employed within probability theory (Billingsley, 1995; Kallenberg, 2002). Alternative theories of conditional probability are available (Dubins, 1975; Popper, 1959; Rényi, 1955), but they have exerted little impact upon mathematical or scientific practice. The main reason is that alternative theories typically impose few quantitative constraints beyond the ratio formula on the relation between conditional and unconditional probabilities, so they offer little useful guidance for computing conditional probabilities. In contrast, the integral formula tightly constrains conditional probabilities in relation to unconditional probabilities. Suppose that C and D are both rcds for P given \mathcal{G} . Then, for every $A \in \mathcal{F}$,

$$C(A|\omega) = D(A|\omega)$$

⁶ A probability measure P is *perfect* iff, for every random variable $X: \Omega \rightarrow \mathbb{R}$, there exists a Borel set $B \subseteq X(\Omega)$ such that $P(X^{-1}(B)) = 1$. If \mathcal{F} is countably generated and P is perfect, then there exists an rcd for P given any sub- σ -field \mathcal{G} (Rao, 2005, pp. 134–135). These conditions are satisfied in virtually all scientific applications of Bayesian decision theory. For further discussion of rcd existence, see (Rao, 2005).

except perhaps for those ω belonging to a set of P -measure 0. Alternative theories of conditional probability usually do not pin down conditional probabilities with nearly so much determinacy. For detailed comparison of rcds with alternative theories, see (Easwaran, 2019).⁷

3 Kolmogorov conditionalization

Rcds figure prominently in many scientific applications of Bayesian decision theory, including within statistics (Florens et al., 1990; Ghosal & van der Vaart, 2017; Schervish, 1995), economics (Feldman, 1987), and cognitive science (Bennett et al., 1996). Something like the following picture underlies these applications. At time t_0 , the agent has *unconditional credences* encoded by a probability space (Ω, \mathcal{F}, P) . She also has *conditional credences* encoded by C , an rcd for P given $\mathcal{G} \subseteq \mathcal{F}$. At a later time t_1 , an exogenous event causes the agent to acquire a new certainty profile δ_ω over \mathcal{G} . Based upon her new certainties over \mathcal{G} , she adopts new credences C_ω over \mathcal{F} . As I will put it, she *uses rcd C to conditionalize on δ_ω* . When an agent uses an rcd to conditionalize, I will say that she engages in *Kolmogorov conditionalization*.⁸

In general, a probability measure P determines conditional probabilities $C(A|\cdot)$ only up to measure 0. Thus, an agent's unconditional credences do not typically determine unique conditional credences C . We must instead take C as an extra primitive element. This is an important difference between Kolmogorov conditionalization and ratio conditionalization, since the ratio formula uniquely determines conditional probabilities $P(H|E)$ when $P(E) > 0$. The extra primitive element seems a small price to pay for the benefits that it buys. Anyway, all theories of conditional probability agree that unconditional probabilities do not uniquely determine conditional probabilities once we move beyond the simple case where $P(E) > 0$.

Kolmogorov conditionalization is a very general update strategy, but it is not universally applicable. There are pathological cases where no rcd exists (Billingsley, 1995, p. 443). Even when an rcd exists, it may not support conditionalization. To see why, say that an update rule C for (Ω, \mathcal{F}) and \mathcal{G} is *proper at ω* iff

$$\text{If } \omega \in G, \text{ then } C(G|\omega) = 1 \quad \text{for all } G \in \mathcal{G}. \quad (8)$$

If C is improper at ω , then there exists $G \in \mathcal{G}$ such that

$$C(G|\omega) < 1 \ \& \ \omega \in G,$$

⁷ As Easwaran (2019) notes, Rényi's (1955) theory may be regarded as generalizing the rcd formalism. Presumably, then, one could reformulate the ideas from the present paper using Rényi's theory. I am not sure whether the reformulation would yield any benefits, but the matter seems worth pursuing. More generally, it would be worth investigating the extent which this paper's ideas could be reformulated using various alternative theories of conditional probability.

⁸ (Rescorla (2018a) proves a Dutch book theorem and converse Dutch book theorem for Kolmogorov conditionalization. The theorems show that Kolmogorov conditionalization is the unique credal update strategy that avoids a sure loss in certain natural learning scenarios. The theorems assume a factive setting. (Rescorla, forthcoming) generalizes the theorems to a non-factive setting. See also (Meehan & Zhang, 2022).

which conflicts with the certainty profile's assignment:

$$\delta_\omega(G) = 1.$$

When such a conflict arises, the agent cannot use C to extend δ_ω to all of \mathcal{F} . Unfortunately, there are cases where every rcd for P given \mathcal{G} is improper at some ω (Blackwell & Dubins, 1975). Fortunately, impropriety occurs rarely if ever in practice. In actual scientific applications, there usually exists an rcd that is proper everywhere.⁹ The probability spaces considered in this paper support rcds that are proper everywhere. For further discussion of impropriety, see (Easwaran, 2011; Meehan & Zhang, 2022; Rescorla, forthcoming).

Let X be a random variable, and let $P_{\sigma(X)}$ be an rcd for P given $\sigma(X)$. Then $P_{\sigma(X)}$ dictates how to update credences in light of newfound certainty that $X = x$. Each possible value x corresponds to a distinct certainty profile that the agent might instantiate. When there are uncountably many possible values x , the model posits uncountably many possible mental states. Some readers may worry that any such model is inapplicable to ordinary humans, since it flouts the apparently finitary nature of human representational and discriminative capacities. The model may seem applicable only to an idealized superhuman with infinitary cognitive abilities that transcend our own.

In evaluating this objection, it is instructive to consider the Bayesian models offered within current cognitive science. Cognitive scientists offer Bayesian models of numerous core mental phenomena (Griffiths et al., 2008), including perception, motor control, decision-making, language acquisition, navigation, social cognition, and causal reasoning. The models have achieved notable explanatory and predictive success, especially as applied to perception (Rescorla, 2015a) and motor control (Rescorla, 2016). Typically, the models posit uncountably many possible mental states. For example, Bayesian perceptual models describe how the perceptual system estimates environmental conditions based on sensory stimulations (Knill & Richards, 1996). The models usually posit uncountably many possible sensory states (e.g. uncountably many possible retinal states) that serve as possible inputs to Bayesian inference. They also posit uncountably many outputs that might result from the Bayesian inference (e.g. uncountably many possible estimates of an object's shape, size, or location). Nevertheless, the models generate powerful psychological explanations (Rescorla, 2018b, 2020). Thus, a Bayesian model may fruitfully apply to ordinary humans even though it posits uncountably many mental states.

I distinguish two possible reactions to such models. The first reaction accepts at face value the postulation of uncountably many mental states. On this reaction, we accept

⁹ The following theorem gives a sufficient condition for propriety almost everywhere (Seidenfeld et al., 2001, p. 1614): *If \mathcal{G} is countably generated, and $P_{\mathcal{G}}$ is an rcd for P given \mathcal{G} , then $P_{\mathcal{G}}$ is proper at ω for P -almost all ω .* One cannot always remove the exceptional set where propriety fails (Blackwell & Ryll-Nardzewski, 1963). To articulate a sufficient condition for propriety everywhere (not just almost everywhere), say that a function $\Phi: \mathcal{F} \rightarrow \mathcal{G}$ is a *selection homomorphism for \mathcal{G} with respect to \mathcal{F}* iff **(a)** it respects complementation and countable union, and **(b)** $\Phi(G) = G$ for every $G \in \mathcal{G}$. The following theorem gives a sufficient condition for existence of an everywhere proper rcd (Sokal, 1981): *If \mathcal{F} is countably generated, and P is perfect, and there exists a selection homomorphism for \mathcal{G} with respect to \mathcal{F} , then there exists an rcd for P given \mathcal{G} that is proper at every ω .* This sufficient condition is satisfied in many cases—including typical applications of Bayesian decision theory.

that an ordinary human can in principle instantiate uncountably many mental states. The second reaction regards the postulation of so many mental states as an *infinitary idealization*, akin to the postulation of an infinitely large biological population within population genetics. On this reaction, a Bayesian model that posits uncountably many mental states should eventually be replaced by a more psychologically realistic model.

Both reactions merit further exploration. The key point for present purposes is that, *on either reaction*, there is a legitimate role for Bayesian models that posit uncountably many mental states. The first reaction holds that such models may be literally true. The second reaction holds that they include infinitary idealizations to be banished from a literally true description. Either way, the mere fact that a Bayesian model posits uncountably many mental states does not bar it from making a useful theoretical contribution. The present paper is offered in that spirit.

4 Certainty eradication

Kolmogorov conditionalization offers a crucial advantage over ratio conditionalization: it can eradicate certainties. This is the flip-side of the fact that Kolmogorov conditionalization can raise probabilities from zero. Here is a simple example (not necessarily involving any kind of epistemic defeat). Suppose that $P(X = x) = 0$ and $P(X \neq x) = 1$. If the agent becomes newly certain that $X = x$, then she can use an rcd $P_{\sigma(X)}$ to conditionalize on a certainty profile δ_ω corresponding to her newfound certainty, i.e. a certainty profile indexed by an ω such that $X(\omega) = x$. She can do so as long as her rcd $P_{\sigma(X)}$ satisfies the condition:

$$P_{\sigma(X)}(X = x|\omega) = 1 \ \& \ P_{\sigma(X)}(X \neq x|\omega) = 0.$$

By using $P_{\sigma(X)}$ to conditionalize, she demotes her former certainty in $X \neq x$ all the down to 0. In general, Kolmogorov conditionalization can raise probabilities from 0 to 1 or anywhere in between, and it can lower probabilities from 1 to 0 or anywhere in between.

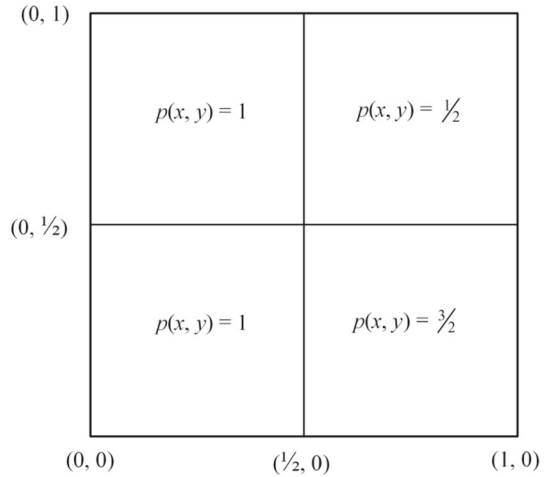
Here is a slightly more elaborate example (still not necessarily involving epistemic defeat). Consider the following *probability density function* $p(x, y)$ over \mathbb{R}^2 :

$$p(x, y) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1/2, 0 \leq y \leq 1 \\ 3/2 & \text{if } 1/2 < x \leq 1, 0 \leq y \leq 1/2 \\ 1/2 & \text{if } 1/2 < x \leq 1, 1/2 < y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

See Fig. 1. For any topological space T , let $\mathcal{B}(T)$ consist of the Borel subsets of T . By integrating $p(x, y)$, we define a probability measure P over $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$:

$$P(H) =_{df} \iint_H p(x, y) dx dy, \text{ for any Borel set } H \subseteq \mathbb{R}^2.$$

Fig. 1 A simple probability density function over \mathbb{R}^2 . All positive probability density is concentrated in the unit square



Suppose an agent has credences given by P . For any x , we have

$$P(X = x \cap Y \leq 1/2) = 0$$

$$P(X \neq x \cup Y > 1/2) = P((X = x \cap Y \leq 1/2)^c) = 1.$$

Let $P_{\sigma(X)}$ be an rcd for P given $\sigma(X)$. Using standard mathematical techniques (Billingsley, 1995, p. 432), one can show that $P_{\sigma(X)}$ must satisfy (9) for almost all (x, y) such that x lies in the unit interval:

$$P_{\sigma(X)}(X = x \cap Y \leq 1/2 | x, y) = \frac{\int_{[0, 1/2]} p(x, t) dt}{\int_{[0, 1]} p(x, t) dt}. \tag{9}$$

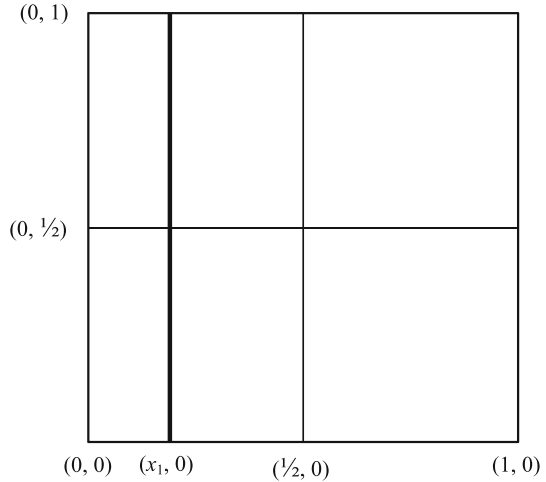
We may choose $P_{\sigma(X)}$ so that it satisfies (9) for *all* (x, y) such that x lies in the unit interval. For all other (x, y) — these lie within an event whose P -measure is 0 — we may choose $P_{\sigma(X)}(\cdot | x, y)$ to be some fixed, arbitrary measure. Pick x_1 such that $0 \leq x_1 \leq 1/2$. Then (9) yields

$$P_{\sigma(X)}(X = x_1 \cap Y \leq 1/2 | x_1, y) = \frac{\int_{[0, 1/2]} dt}{\int_{[0, 1]} dt} = 1/2, \tag{10}$$

for any y . See Fig. 2. An agent who becomes certain that $X = x_1$ and who updates her credences using $P_{\sigma(X)}$ will raise her credence in $X = x_1 \cap Y \leq 1/2$ from 0 to $1/2$. This is the intuitively correct reaction: the agent concentrates all credal mass on the vertical line $X = x_1$ from 0 to 1, where the prior probability density $p(x, y)$ is constant. Since $P_{\sigma(X)}(\cdot | x_1, y)$ is a probability measure, (10) entails

$$P_{\sigma(X)}(X \neq x_1 \cup Y > 1/2 | x_1, y) = 1/2. \tag{11}$$

Fig. 2 Assuming $0 \leq x_1 \leq 1/2$, probability density $p(x, y)$ is constant along the vertical line $X = x_1$. If the agent conditionalizes based on newfound certainty that $X = x_1$, then she will allocate all her credence uniformly over this vertical line



Thus, our agent demotes her former certainty in $X \neq x_1 \cup Y > 1/2$ from 1 to $1/2$. In contrast, pick x_2 such that $1/2 < x_2 \leq 1$. Then (9) yields

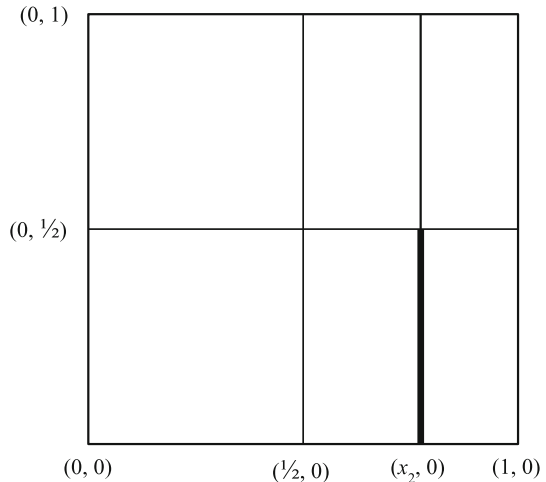
$$P_{\sigma(X)}(X = x_2 \cap Y \leq 1/2 | x_2, y) = \frac{\int_{[0, 1/2]} 3/2 dt}{\int_{[0, 1/2]} 3/2 dt + \int_{(1/2, 1]} 1/2 dt} = \frac{3/4}{3/4 + 1/4} = 3/4, \quad (12)$$

which entails

$$P_{\sigma(X)}(X \neq x_2 \cup Y > 1/2 | x_2, y) = 1/4. \quad (13)$$

See Fig. 3. An agent who become certain that $X = x_2$ and who updates her credences using $P_{\sigma(X)}$ will raise her credence in $X = x_2 \cap Y \leq 1/2$ from 0 to $3/4$, and she will lower

Fig. 3 Assuming $1/2 < x_2 \leq 1$, probability density $p(x, y)$ is higher along the bottom half of the vertical line $X = x_2$ than along the top half. If the agent conditionalizes based on newfound certainty that $X = x_2$, then she will allocate all her credence over this vertical line, with more credal mass assigned to the bottom half than the top half



her credence in $X \neq x_2 \cup Y > \frac{1}{2}$ in from 1 to $\frac{1}{4}$. Again, this makes intuitive sense: the agent now concentrates all credal mass on the vertical line $X = x_2$ from 0 to 1, and $p(x, y)$ is weighted three times higher along the bottom half of the line.

The foregoing observations prompt us to reflect upon the meaning of “certainty” in Bayesian decision theory. “Certainty” may seem to connote immutable confidence that a proposition is true. Yet that is *not* what “certainty” means—not if we define “certainty” as “assignment of credence 1.” As noted in Sect. 2, the probability calculus axioms entail that $P(X = x) = 0$ for all but countably many values x of a random variable X . When you set $P(X \neq x) = 1$ and $P(X = x) = 0$, it does not follow that you regard $X = x$ as metaphysically impossible, or that have definitively ruled out $X = x$, or that no possible evidence could lead you to assign non-zero credence to $X = x$. All that follows is that you regard $X = x$ as vanishingly unlikely. The probability calculus axioms entail that you *must* regard $X = x$ as vanishingly unlikely for all but countably many values x . Even though your credence in $X = x$ is currently 0, you are prepared to raise this credence in light of new evidence. If you do so, you must simultaneously downgrade your certainty in $X \neq x$. You will typically downgrade many other certainties as well, as illustrated by (11) and (13). Kolmogorov conditionalization provides a principled basis for these credal transitions.

5 Certainty gained, then lost

Since Kolmogorov conditionalization can eradicate certainties, it is much more flexible than ratio conditionalization. In what follows, I leverage the increased flexibility to model defeasible inference across a range of cases. This section warms up by modeling an example where an agent conditionalizes on a proposition and then loses certainty in the proposition. Defects in the model will motivate refinements made in Sect. 6.

At time t_0 , John awaits his medical test result for a rare disease. At time t_1 , he receives his test result. Upon reading the report, he becomes certain that the test result was positive. He conditionalizes on the positive test result, substantially raising his credence that he has the disease. At time t_2 , John re-reads the medical report and realizes that he misinterpreted it. In fact, the test result was negative. Intuitively, he should now conditionalize on the negative test result and downgrade his credence that he has the disease. But he cannot do so using ratio conditionalization because at t_1 he assigned zero credence to the negative test result. Call this example **False Alarm**.

Using rcds, we can elaborate False Alarm into a model that includes certainty eradication at t_2 . Assume a suitable outcome space Ω , and let

- Disease* = the set of outcomes in which John has the rare disease.
- No Disease* = the set of outcomes in which John does not have the rare disease.
- Positive* = the set of outcomes in which the test has a positive result.
- Negative* = the set of outcomes in which the test has a negative result.

I assume that the test can only have a positive or negative result, so that *Positive* and *Negative* are complements. Let \mathcal{F} be the σ -field generated by *Disease* and *Positive*. John’s credences at t_0 are given by (Ω, \mathcal{F}, P) , where P is the unique probability measure over \mathcal{F} such that:

$$\begin{aligned}
 P(Disease) &= .025 & P(No\ Disease) &= .975 \\
 P(Positive|Disease) &= .95 & P(Positive|No\ Disease) &= .05 \\
 P(Negative|Disease) &= .05 & P(Negative|No\ Disease) &= .95
 \end{aligned}$$

These credences reflect the base rate for the disease, the frequency of false positives, and the frequency of false negatives. By the law of total probability,

$$\begin{aligned}
 P(Positive) &= P(Positive|Disease)P(Disease) + P(Positive|No\ Disease)P(No\ Disease) \\
 &= .95 \times .025 + .05 \times .975 = .0725,
 \end{aligned}$$

so that $P(Negative) = .9275$. By Bayes's theorem,

$$P(Disease|Positive) = \frac{P(Disease)P(Positive|Disease)}{P(Positive)} = \frac{.025 \times .95}{.0725} = 19/58$$

$$P(Disease|Negative) = \frac{P(Disease)P(Negative|Disease)}{P(Negative)} = \frac{.025 \times .05}{.9275} = 1/742$$

At t_1 , John conditionalizes on *Positive*, acquiring new credences

$$\begin{aligned}
 P_{t_1}(Positive) &= 1 \\
 P_{t_1}(Negative) &= 0 \\
 P_{t_1}(Disease) &= 19/58
 \end{aligned}$$

So far, so standard.

Now comes the less standard part. To model John's credal transition at t_2 , let \mathcal{G} be the sub- σ -field

$$\{0, Positive, Negative, \Omega\}.$$

John's realization that the test was negative corresponds to a certainty profile δ_ν , where ν is any outcome belonging to *Negative*. We model John's conditional credences at t_1 by stipulating

$$\begin{aligned}
 \mu(Disease|\omega) &= \begin{cases} 19/58 & \text{if } \omega \in Positive \\ 1/742 & \text{if } \omega \in Negative \end{cases} \\
 \mu(Positive|\omega) &= \begin{cases} 1 & \text{if } \omega \in Positive \\ 0 & \text{if } \omega \in Negative \end{cases}
 \end{aligned}$$

These stipulations extend to a unique function $\mu: \mathcal{F} \times \Omega \rightarrow \mathbb{R}$ such that μ_ω is a probability measure for each $\omega \in \Omega$. It is easy to check that μ is an rcd for P_{t_1} given \mathcal{G} . Specifically, the integral formula is trivially satisfied. At t_2 , John uses μ to conditionalize on δ_ν , where ν is any outcome belonging to *Negative*. Thus, he acquires

new credences

$$\begin{aligned}P_{t_2}(\textit{Positive}) &= 0 \\P_{t_2}(\textit{Negative}) &= 1 \\P_{t_2}(\textit{Disease}) &= 1/742\end{aligned}$$

He downgrades his certainty in *Positive* all the way down to 0 and downgrades his credence in *Disease* from 19/58 to 1/742.

Our model illustrates the increased flexibility afforded by Kolmogorov conditionalization. Over the course of John's credal evolution, his credence in *Positive* goes from .0725 to 1 to 0, and his credence in *Disease* changes as dictated by his conditional credences. The postulated credal transitions look quite reasonable, given John's initial credences.

Nevertheless, there is something disturbingly trivial about the model. I stipulated that $\mu(\textit{Disease}|\omega) = 1/742$ if $\omega \in \textit{Negative}$, but I could have stipulated $\mu(\textit{Disease}|\omega) = .5$, or .9999, or even 1. Each alternative stipulation would also yield an rcd for P_{t_1} given \mathcal{G} . Thus, John's unconditional credences at t_1 leave his conditional credences at t_1 completely undetermined. Any alternative update rule μ^* would satisfy the integral formula just as well as μ . Such extreme flexibility is undesirable. Ideally, a final theory of rational inference will pin down more determinately how credences evolve over time.

Our model of False Alarm shows that Kolmogorov conditionalization can, in principle, support acquisition and loss of certainties. But the model taken on its own is unsatisfying because it hinges upon arbitrary stipulation of an rcd μ . We must try to do better.

6 Rigidity

I now advance a rational norm, inspired by Jeffrey (1983), that tightly constrains credal evolution. I call the norm *Rigidity*. Section 6.1 introduces Rigidity. Section 6.2 discusses how Rigidity yields an improved treatment of False Alarm and similar examples.

6.1 Minimal change in conditional probabilities

Consider again the situation emphasized by Jeffrey: an external influence causes an agent to reallocate credences across a partition $\mathcal{E} = \{E_i\}$, and on that basis the agent must assign credences to all remaining propositions. Why should we accept Jeffrey's recommended credal update strategy (2)? Beginning with Jeffrey (1983), and continuing through the later literature (e.g. Earman, 1992, pp. 34–35; Joyce, 2009, pp. 35–35; Weisberg, 2009b), philosophers often motivate (2) by citing the invariance condition

$$P_{old}(H|E_i) = P_{new}(H|E_i) \quad \text{for all } H \text{ and all } E_i \text{ such that } P_{new}(E_i) > 0. \quad (14)$$

(2) follows from (14) together with the law of total probability (7).

The rationale underlying (14) is that, when all credal change stems from reallocation across a partition, probabilities conditional on partition propositions should remain fixed. Intuitively: reallocating probabilities *across* a partition tells you nothing new about how probability mass should be distributed *inside* any member of the partition. Credal reallocation over the partition provides no rational basis for changing your credences *conditional on a given partition proposition*. As Joyce (2009, p. 36) notes, this rationale reflects “a kind of minimal change ‘ethos’ which prohibits the posterior from introducing distinctions in probability among hypotheses that are not already inherent in the prior or explicitly mandated by new evidence.” (14) enforces the minimal change ethos by holding conditional probabilities as fixed as possible, given that probabilities assigned to partition propositions have changed.

The rationale for (14) has a causal dimension: we assume that an external event triggers the transition from P_{old} to P_{new} by instilling new credences across a partition \mathcal{E} . Different authors express this causal assumption in different ways. Earman (1992, p. 34) says that credal changes are “generated” by new credences across the partition. Joyce (2009, pp. 35–36) posits an event whose “only immediate effect” is to fix new credences for partition propositions. In (Rescorla, 2021), I said that the new credal assignment over the partition “mediates” the transition from P_{old} to P_{new} . The core idea behind these varying formulations is that an event alters the agent’s credences *entirely by way of* altering her credences over \mathcal{E} . Intuitively: we restrict attention to situations where all credal change stems from the new credal assignment over \mathcal{E} . Virtually all discussions of Jeffrey Conditionalization assume a restriction along these lines, although the restriction often figures only implicitly.

An important task for formal epistemology is to analyze more systematically the assumed restriction on causal structure. In (Rescorla, 2021), I offered one possible analysis. But I will not assume that analysis, or any other analysis. Even lacking a detailed analysis, the basic idea seems clear enough for present purposes.

It will prove helpful to articulate a more precise statement of the diachronic credal norm corresponding to (14). In (Rescorla, 2021), I formulated the norm as imposing the following requirement:

If an agent begins with credences P_{old} , and $\mathcal{E} = \{E_i\}$ is a countable set of mutually exclusive, jointly exhaustive propositions such that $P_{old}(E_i) > 0$ for each i , and she subsequently adopts new credences P_{new} such that

$$\sum_i P_{new}(E_i) = 1, \text{ and the new credal assignment over } \mathcal{E} \text{ mediates the transition}$$

from P_{old} to P_{new} , then $P_{old}(\cdot|E_i) = P_{new}(\cdot|E_i)$ for all i such that $P_{new}(E_i) > 0$,

(15)

where the clause “the new credal assignment over \mathcal{E} mediates the transition from P_{old} to P_{new} ” reflects a causal assumption that credences change *solely* due to the new credal assignment over \mathcal{E} . For present purposes, one could equally well express the causal assumption through the language used by Earman (“generated”) or Joyce (“only immediate effect”).

The antecedent of (15) confines attention to situations where $P_{old}(E_i) > 0$ for each i , and the consequent addresses only those E_i such that $P_{new}(E_i) > 0$. These restrictions ensure that conditional probabilities, as specified by the ratio formula, are well-defined. However, I see no reason to impose the restrictions once we have in hand conditional probabilities beyond those given by the ratio formula. The intuitive rationale supplied by Joyce's minimal change ethos applies just as well to cases where $P_{old}(E_i) = 0$ or $P_{new}(E_i) = 0$.

Indeed, the rationale applies just as well to numerous cases that do not feature a countable partition \mathcal{E} . Consider an agent with credences modeled by a probability space (Ω, \mathcal{F}, P) , and let $\mathcal{G} \subseteq \mathcal{F}$ be a sub- σ -field. \mathcal{G} may or may not be generated by a countable partition. Suppose that the agent has conditional credences over \mathcal{F} given \mathcal{G} , modeled by update rule C . Suppose that there occurs an exogenous shift in credal mass across \mathcal{G} , inducing further credal changes across \mathcal{F} . Following Meehan and Zhang (2020), I propose that the agent's conditional credences C should remain fixed. More carefully, credal transitions should satisfy the following requirement:

RIGIDITY: If an agent begins with unconditional credences P_{old} over \mathcal{F} and conditional credences C_{old} over \mathcal{F} given \mathcal{G} , and she subsequently adopts new credences P_{new} over \mathcal{F} and new conditional credences C_{new} over \mathcal{F} given \mathcal{G} , and the new credal assignment over \mathcal{G} mediates the transition from P_{old} and C_{old} to P_{new} and C_{new} , then $C_{old} = C_{new}$,

where the clause "the new credal assignment over \mathcal{G} mediates the transition from P_{old} and C_{old} to P_{new} and C_{new} " reflects a causal assumption that the agent's conditional and unconditional credences change *solely* due to the new credal assignment over \mathcal{G} . Again, we could equally well substitute other locutions that express the same causal assumption. Like (15), Rigidity is a minimal change principle. It leaves conditional probabilities as fixed as possible, given that credences over \mathcal{G} have changed. Intuitively: credal reallocation *across* a sub- σ -field provides no rational basis for changes in credence *conditional on the sub- σ -field*.

It would be good to explore more fully the basis for Rigidity. In this paper, I focus on *applying* Rigidity. Actually, I will apply a fairly weak consequence of Rigidity:

WEAK RIGIDITY: If an agent begins with credences P_{old} over \mathcal{F} and conditional credences C_{old} over \mathcal{F} given \mathcal{G} , and she subsequently adopts new credences P_{new} over \mathcal{F} and new conditional credences C_{new} over \mathcal{F} given \mathcal{G} , and $P_{new}|\mathcal{G} = \delta_\omega$ for some ω , and the new credal assignment over \mathcal{G} mediates the transition from P_{old} and C_{old} to P_{new} and C_{new} , then $C_{old} = C_{new}$,

where $P_{new}|\mathcal{G}$ is the restriction of P_{new} to \mathcal{G} . Weak Rigidity confines attention to cases where the new credal assignment over \mathcal{G} is a certainty profile. For discussion of cases where the new credal assignment over \mathcal{G} is not a certainty profile, see (Meehan & Zhang, 2020).

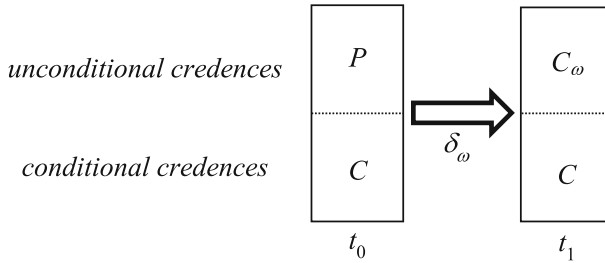


Fig. 4 The agent gains new certainty profile δ_ω and responds by using C to conditionalize on δ_ω . Complying with Weak Rigidity, she retains her conditional credences C . In this diagram, and in subsequent such diagrams, unconditional credences lie above the dotted line and conditional credences lie below the dotted line

6.2 Credal evolution conforming to Rigidity

I propose the following picture of credal evolution. At t_0 , an agent has unconditional credences modeled by a probability space (Ω, \mathcal{F}, P) and conditional credences modeled by C , an rcd for P given \mathcal{G} . At t_1 , an exogenous change instills certainty profile δ_ω over \mathcal{G} . Assume that C is proper at ω . Then the agent can use C to conditionalize on δ_ω , adopting C_ω as her new credal allocation over \mathcal{F} . Her unconditional credences at t_1 are modeled by $(\Omega, \mathcal{F}, C_\omega)$. Complying with Rigidity, she retains her conditional credences C . See Fig. 4.

Theorem *Let (Ω, \mathcal{F}, P) be a probability space, let $\mathcal{G} \subseteq \mathcal{F}$ be a sub- σ -field, and let C be an rcd for P given \mathcal{G} . Suppose that C is proper at ω . Then C is an rcd for C_ω given \mathcal{G} .*

Proof The only non-trivial clause is the integral formula, with C_ω serving as the unconditional probability measure: we must show that

$$C_\omega(A \cap G) = \int_G C(A|v) dC_\omega(v), \tag{16}$$

for every $A \in \mathcal{F}$ and $G \in \mathcal{G}$. Fix $A \in \mathcal{F}$ and $G \in \mathcal{G}$. Define

$$H =_{df} \{v \in \Omega : C(A|\omega) = C(A|v)\}.$$

$C(A|\cdot): \Omega \rightarrow \mathbb{R}$ is \mathcal{G} -measurable, and H is the inverse image of $\{C(A, \omega)\}$ under $C(A|\cdot)$, so $H \in \mathcal{G}$. Since $\omega \in H$, it follows from (8) that

$$C_\omega(H) = 1 \ \& \ C_\omega(H^c) = 0.$$

We now calculate:

$$\begin{aligned} \int_G C(A|v) dC_\omega(v) &= \int_{G \cap H} C(A|v) dC_\omega(v) + \int_{G \cap H^c} C(A|v) dC_\omega(v) \\ &= \int_{G \cap H} C(A|\omega) dC_\omega(v) + 0 = C_\omega(A) \int_{G \cap H} dC_\omega(v) = C_\omega(A) C_\omega(G \cap H). \end{aligned}$$

Either $\omega \in G$ or $\omega \notin G$. If $\omega \in G$, then (8) entails

$$C_\omega(G) = 1,$$

so that

$$C_\omega(A)C_\omega(G \cap H) = C_\omega(A) = C_\omega(A \cap G),$$

which confirms (16). If $\omega \notin G$, then (8) entails.

$$C_\omega(G) = 0,$$

so that

$$C_\omega(A)C_\omega(G \cap H) = 0 = C_\omega(A \cap G),$$

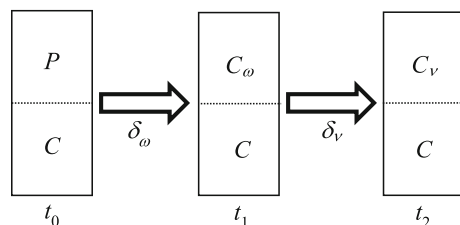
which also confirms (16). We have therefore shown that C is an rcd for C_ω given \mathcal{G} . □

Consider again the agent depicted in Fig. 4: she begins with unconditional credences (Ω, \mathcal{F}, P) and conditional credences modeled by C , then transitions based on certainty profile δ_ω to new unconditional credences C_ω while holding fixed her conditional credences C . We have assumed that C is proper at ω . (Otherwise, the agent cannot use C to conditionalize on δ_ω in the first place.) Our theorem entails that C is also an rcd for the agent’s new unconditional credal allocation C_ω . The agent may therefore continue to use C to conditionalize. If a new exogenous change instills a new certainty profile δ_ν at t_2 , then she can use C to conditionalize on the new certainty profile δ_ν , so long as C is proper at ν . See Fig. 5.

Figure 5 is quite general. It applies to a wide range of situations in which an agent gains a certainty profile δ_ω over a sub- σ -field and then gains a different certainty profile δ_ν over the same sub- σ -field. I do not say Fig. 5 applies to all such situations. If an rcd does not exist, or if every rcd is improper at ω or ν , then Fig. 5 does not apply. However, such situations arise rarely if ever in scientific applications.

To illustrate the virtues of Fig. 5, let us revisit False Alarm. We saw in Sect. 5 that John can use μ at t_2 to conditionalize on his newfound certainty in *Negative*. The worry raised in Sect. 4.2 was that μ seemed arbitrary. Why should John update his credences using μ rather than another rcd μ^* ? Rigidity enables a principled answer. The key point here is that μ encodes conditional probabilities that John has at t_0 . The conditional probabilities follow from our choice of P and from the ratio formula.

Fig. 5 The agent gains new certainty profile δ_ω and responds by using C to conditionalize on δ_ω . Then she gains new certainty profile δ_ν and responds by using C to conditionalize on δ_ν



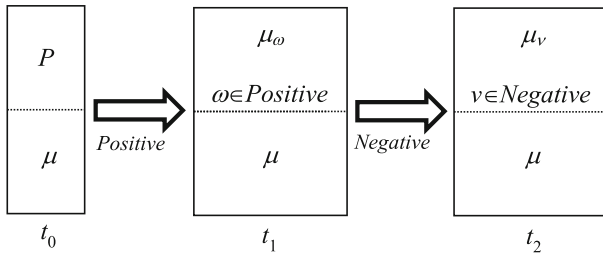


Fig. 6 Credal evolution in False Alarm. John’s initial unconditional credences P determine his conditional credences μ , via the ratio formula. John becomes certain of *Positive* and responds by forming new credences μ_ω , where ω is any arbitrary outcome belonging to *Positive*. Then he becomes certain of *Negative* and responds by forming new credences μ_ν , where ν is any arbitrary outcome belonging to *Negative*. Complying with Weak Rigidity, he retains his conditional credences μ throughout this process

When John conditionalizes on *Positive* at t_1 , Rigidity mandates that he leave those conditional probabilities fixed. John’s fixed conditional probabilities, codified by μ , serve as a basis for conditionalization when at t_2 he becomes certain of *Negative*. Hence, Rigidity rationalizes the choice of μ rather than any alternative rcd μ^* . Given John’s unconditional credences at t_0 , μ is the unique rcd that conforms to Rigidity. See Fig. 6.

Our revised analysis of False Alarm illustrates the benefits that Rigidity offers to Kolmogorov conditionalizers. By constraining conditional credence, Rigidity guides the course of iterated Bayesian inference. Here is another example along the same lines:

Mismeasurement Jane is a scientist with credences P at time t_0 . X is a random variable that reflects the outcome of an experiment. At t_1 , Jane becomes certain that $X = x_1$ and updates her other credences on that basis. At t_2 , she realizes that her certainty in $X = x_1$ was misplaced: she misread a measuring instrument, or the measuring instrument was poorly calibrated, or she was deceived by her assistant, etc. Jane becomes newly certain that $X = x_2$. How should she proceed?

Using Rigidity, we can elaborate Mismeasurement so as to include a principled credal update at t_2 . Stipulate that Jane has conditional credences at t_0 given by $P_{\sigma(X)}$, an rcd for P given $\sigma(X)$. At t_1 , she becomes certain that $X = x_1$ and on that basis acquires new credences $P_{\sigma(X)}(\cdot|\omega_1)$, where ω_1 is any outcome such that $X(\omega_1) = x_1$. Complying with Rigidity, she carries her conditional credences $P_{\sigma(X)}$ forward from t_0 to t_1 . At t_2 , she becomes certain that $X = x_2$ and on that basis acquires new credences $P_{\sigma(X)}(\cdot|\omega_2)$, where ω_2 is any outcome such that $X(\omega_2) = x_2$. Given Rigidity, her conditional credences at t_0 uniquely determine how she should update her credences upon becoming certain at t_1 that $X = x_1$ and also how she should update her credences upon becoming certain at t_2 that $X = x_2$. See Fig. 7. Jane can implement Fig. 7 as long as $P_{\sigma(X)}$ is proper at ω_1 and ω_2 .

More specifically, suppose that at t_0 Jane has the probability density $p(x, y)$ given by Fig. 1. This is not a useful density for real-world applications, but it suffices for heuristic purposes. At t_1 , Jane becomes certain that $X = x_1$, where $0 \leq x_1 \leq 1/2$. She forms new credences $P_{\sigma(X)}(\cdot|x_1, y)$, where y is any real number. In particular, she forms

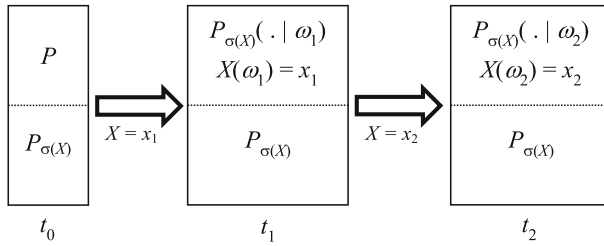


Fig. 7 Credal evolution in Mismeasurement

the credences given by (10) and (11). At t_2 , she becomes certain that $X = x_2$, where $\frac{1}{2} < x_2 \leq 1$. She forms new credences $P_{\sigma(X)}(\cdot | x_2, y)$, including the credences given by (12) and (13). Over the course of her credal evolution, her credence in $X \neq x_1 \cup Y > \frac{1}{2}$ goes from 1 to $\frac{1}{2}$ to $\frac{3}{4}$.

False Alarm and Mismeasurement illustrate how Rigidity can steer Kolmogorov conditionalizers through rational acquisition and loss of certainties. In each example, the agent’s initial doxastic state dictates how she should reallocate credence when she becomes certain of a conditioning proposition *and also* when she later becomes certain of a conflicting proposition.

Of course, we must assume initial conditional and unconditional credences to derive a determinate credal reallocation policy. But this is no problem for my approach, because *all* Bayesian theorizing assumes that the agent has certain initial credences. One must always assume some credal starting point. As noted in Sect. 2, the assumed credal starting point will usually include primitive *conditional* credences over and above the agent’s unconditional credences once we move beyond simple cases where the ratio formula prevails. The essence of the Bayesian framework is to place rational constraints on credal evolution *given the agent’s initial conditional and unconditional credences*. That is precisely what Rigidity accomplishes.

Figure 5 lends itself to iteration. Consider an agent who transitions from certainty profile δ_{ω_1} to certainty profile δ_{ω_2} to certainty profile δ_{ω_3} , and so on, all over a fixed sub- σ -field \mathcal{G} . So long as C is proper at each index $\omega_1, \omega_2, \omega_3, \dots, \omega_n, \dots$ the agent can carry C forward at each stage, using it as her fixed update rule. At each stage, the agent downgrades certainties acquired at the previous stage. See Fig. 8.

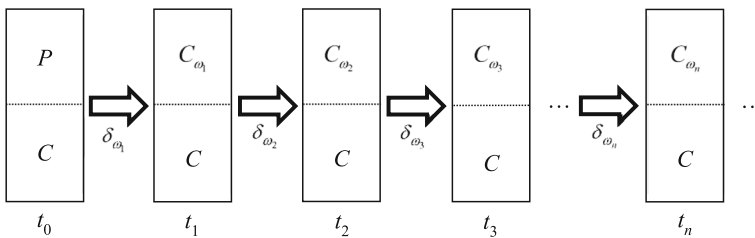


Fig. 8 Sequential credal updates conforming to Weak Rigidity. At each time t_n after the starting point t_0 , the agent gains new certainty profile δ_{ω_n} and responds by using C to conditionalize on δ_{ω_n} , yielding new credences C_{ω_n} . Complying with Weak Rigidity, she retains her conditional credences C

7 Generalized rigidity

Rigidity applies to situations where credences shift over a single fixed conditioning sub- σ -field \mathcal{G} . In many applications, though, there is not a single fixed conditioning sub- σ -field. Instead, the agent accumulates new evidence over a non-decreasing sequence of sub- σ -fields:

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{G}_3 \dots \subseteq \mathcal{G}_n \subseteq \dots$$

For example, an agent might progressively learn the values of random variables $X_1, X_2, \dots, X_n, \dots$, a situation we can model using the sub- σ -fields:

$$\sigma(X_1) \subseteq \sigma(X_1, X_2) \subseteq \sigma(X_1, X_2, X_3) \subseteq \dots \subseteq \sigma(X_1, X_2, \dots, X_n) \subseteq \dots$$

I will now generalize Rigidity so as to accommodate sequential evidence accumulation.

7.1 Minimal change revisited

Consider an agent who begins with unconditional credences modeled by a probability space (Ω, \mathcal{F}, P) . Fix \mathcal{G}_1 and \mathcal{G}_2 such that $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{F}$. Suppose that the agent has conditional credences over \mathcal{F} given \mathcal{G}_2 , modeled by an update rule $C: \mathcal{F} \times \Omega \rightarrow \mathbb{R}$. Suppose that there is an exogenous shift in credences across \mathcal{G}_1 , inducing additional credal changes over the rest of \mathcal{F} . I submit that this shift should leave fixed the agent’s credences conditional on \mathcal{G}_2 . More precisely, I propose the following requirement on credal evolution:

GENERALIZED RIGIDITY: If an agent begins with credences P_{old} over \mathcal{F} and conditional credences C_{old} over \mathcal{F} given \mathcal{G}_2 , and she subsequently adopts new credences P_{new} over \mathcal{F} and new conditional credences C_{new} over \mathcal{F} given \mathcal{G}_2 , and $\mathcal{G}_1 \subseteq \mathcal{G}_2$, and the new credal assignment over \mathcal{G}_1 mediates the transition from P_{old} and C_{old} to P_{new} and C_{new} , then $C_{old} = C_{new}$,

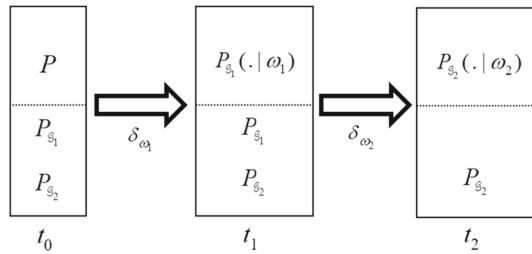
where the clause “the new credal assignment over \mathcal{G}_1 mediates the transition from P_{old} and C_{old} to P_{new} and C_{new} ” registers that the agent’s credences change *solely* due to the new credal assignment over \mathcal{G}_1 . Generalized Rigidity demands that, in such situations, credences conditional on \mathcal{G}_2 remain constant. Note that Generalized Rigidity entails Rigidity. Like Rigidity, Generalized Rigidity is a minimal change principle. Intuitively: credal reallocation across a sub- σ -field provides no basis for changing credences *conditional on a larger sub- σ -field*.

To illustrate, consider a partition $\mathcal{E} = \{E_i\}$ and a finer-grained partition $\mathcal{D} = \{E_{ij}\}$, where

$$E_i = \bigcup_j E_{ij}.$$

Suppose that the agent at time t_1 gains new credences over \mathcal{E} and that these new credences cause her to reallocate credences over all remaining propositions. Intuitively,

Fig. 9 Two credal updates conforming to Weak Generalized Rigidity. Note that the agent no longer has conditional credences $P_{\mathcal{G}_1}$ at t_2



the agent’s new credal allocation over \mathcal{E} should not lead her to change her probabilities conditional on members of \mathcal{D} . Reallocating credence over the E_i should not affect how credence is allocated within each E_i , so it should not change how credence is allocated within each E_{ij} . Indeed, one can easily show that Jeffrey conditionalization in response to new credences over \mathcal{E} leaves fixed all conditional probabilities $P(H|E_{ij})$. Let $\mathcal{G}_1 = \sigma(\mathcal{E})$, the σ -field generated by \mathcal{E} , and let $\mathcal{G}_2 = \sigma(\mathcal{D})$, the σ -field generated by \mathcal{D} . Note that $\mathcal{G}_1 \subseteq \mathcal{G}_2$. If the agent updates using Jeffrey Conditionalization, she will conform to Generalized Rigidity as applied to \mathcal{G}_1 and \mathcal{G}_2 .

As with the original version of Rigidity, I focus exclusively on a weak consequence of Generalized Rigidity:

WEAK GENERALIZED RIGIDITY: If an agent begins with credences P_{old} over \mathcal{F} and conditional credences C_{old} over \mathcal{F} given \mathcal{G}_2 , and she subsequently adopts new credences P_{new} over \mathcal{F} and new conditional credences C_{new} over \mathcal{F} given \mathcal{G}_2 , and $\mathcal{G}_1 \subseteq \mathcal{G}_2$, and $P_{new}|\mathcal{G}_1 = \delta_\omega$ for some ω , and the new credal assignment over \mathcal{G}_1 mediates the transition from P_{old} and C_{old} to P_{new} and C_{new} , then $C_{old} = C_{new}$.

Weak Generalized Rigidity entails Weak Rigidity. I will use Weak Generalized Rigidity to model cases of defeasible inference that cannot be modeled using Weak Rigidity.

7.2 Credal evolution conforming to generalized rigidity

I propose the following picture of credal evolution. At t_0 , an agent has unconditional credences modeled by a probability space (Ω, \mathcal{F}, P) . She also has conditional credences modeled by $P_{\mathcal{G}_1}$, a proper rcd for P given \mathcal{G}_1 , and $P_{\mathcal{G}_2}$, a proper rcd for P given \mathcal{G}_2 , where $\mathcal{G}_1 \subseteq \mathcal{G}_2$. Moreover, for each $\omega \in \Omega$, $P_{\mathcal{G}_2}$ is an rcd for $P_{\mathcal{G}_1}(\cdot | \omega)$ given \mathcal{G}_2 . At t_1 , an exogenous change instills certainty profile δ_{ω_1} over \mathcal{G}_1 . In response, the agent uses $P_{\mathcal{G}_1}$ to conditionalize on δ_{ω_1} , adopting $P_{\mathcal{G}_1}(\cdot | \omega_1)$ as her new credal allocation over \mathcal{F} . The agent retains the same fixed conditional credences $P_{\mathcal{G}_1}$ and $P_{\mathcal{G}_2}$, as mandated by Generalized Rigidity. At t_2 , an exogenous change instills certainty profile δ_{ω_2} over \mathcal{G}_2 . In response, the agent uses $P_{\mathcal{G}_2}$ to conditionalize on δ_{ω_2} , adopting $P_{\mathcal{G}_2}(\cdot | \omega_2)$ as her new credal allocation over \mathcal{F} . Her credences conditional on \mathcal{G}_2 are still given by $P_{\mathcal{G}_2}$, as Generalized Rigidity mandates. See Fig. 9.

Note that, at t_2 , Generalized Rigidity does *not* require the agent’s credences conditional on \mathcal{G}_1 to be given by $P_{\mathcal{G}_1}$. This is as it should be. To see why, fix events E and F such that $P(E \cap F) > 0$. If the agent ratio conditionalizes on $E \cap F$, then she will not

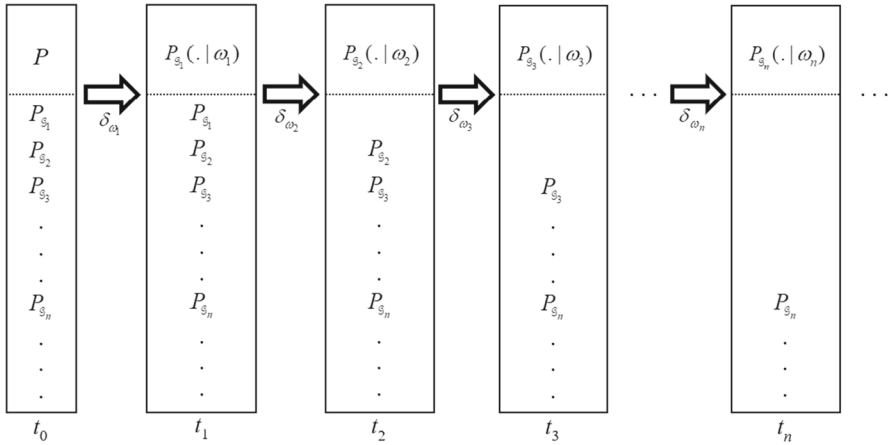


Fig. 10 Sequential credal updates conforming to Weak Generalized Rigidity. At each time t_n after the initial starting point t_0 , the agent uses $P_{\mathcal{G}_n}$ to conditionalize on δ_{ω_n} , adopting new credences $P_{\mathcal{G}_n}(\cdot | \omega_n)$. Note that the agent no longer has conditional credences $P_{\mathcal{G}_{n-1}}$ at time t_n

usually retain her initial conditional probabilities $P(H|E)$. She is now certain of E , so her new credence in H conditional on E is simply her new unconditional credence in H :

$$P_{new}(H|E) = P_{new}(H) = P(H|E \cap F),$$

which may differ from $P(H|E)$. Intuitively: newfound certainty in F may alter credences conditional on E . Restating the point using the rcd formalism, let \mathcal{G}_1 be the σ -field generated by $\{E\}$, and let \mathcal{G}_2 be the σ -field generated by $\{E, F\}$. Assume that $P_{\mathcal{G}_1}$ is an rcd for P given \mathcal{G}_1 and that $P_{\mathcal{G}_2}$ is an rcd for P given \mathcal{G}_2 , where these rcds are given by the ratio formula. If $\omega_2 \in E \cap F$, then $P_{\mathcal{G}_1}$ may not be an rcd for $P_{\mathcal{G}_2}(\cdot | \omega_2)$ given \mathcal{G}_1 . Intuitively: newfound certainties regarding \mathcal{G}_2 may alter probabilities conditional on \mathcal{G}_1 .

We may extend Fig. 9 to scenarios where the agent sequentially accumulates evidence at times $t_1, t_2, \dots, t_n, \dots$ regarding sub- σ -fields

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{G}_3 \dots \subseteq \mathcal{G}_n \subseteq \dots$$

Suppose that $P_{\mathcal{G}_n}$ is a proper rcd for P given \mathcal{G}_n . Suppose also that, for each $\omega \in \Omega$, $P_{\mathcal{G}_{n+1}}$ is an rcd for $P_{\mathcal{G}_n}(\cdot | \omega)$ given \mathcal{G}_{n+1} . At t_n , an exogenous change instills certainty profile δ_{ω_n} over \mathcal{G}_n . In response, the agent uses $P_{\mathcal{G}_n}$ to conditionalize on δ_{ω_n} , adopting $P_{\mathcal{G}_n}(\cdot | \omega_n)$ as her new credal allocation over \mathcal{F} . Complying with Generalized Rigidity, she retains the conditional credences given by $P_{\mathcal{G}_n}, P_{\mathcal{G}_{n+1}}, P_{\mathcal{G}_{n+2}}, \dots, P_{\mathcal{G}_{n+m}}, \dots$. She uses those conditional credences for credal updates at $t_{n+1}, \dots, t_{n+m}, \dots$. In this manner, the agent’s initial conditional credences dictate her credal evolution as she sequentially gains new evidence. See Fig. 10.

In presenting Figs. 9 and 10, I made two assumptions about the rcds $P_{\mathcal{G}_n}$. First, $P_{\mathcal{G}_n}$ is proper at ω_n . Second, $P_{\mathcal{G}_{n+1}}$ is an rcd for $P_{\mathcal{G}_n}(\cdot|\omega_n)$ given \mathcal{G}_{n+1} . Only under those assumptions can a Kolmogorov conditionalizer use the rcds $P_{\mathcal{G}_n}$ to conditionalize at each time stage t_n .

Unfortunately, there is no global guarantee that the two assumptions are satisfiable. That $P_{\mathcal{G}_{n+1}}$ is an rcd for P given \mathcal{G}_{n+1} does not guarantee that it is an rcd for $P_{\mathcal{G}_n}(\cdot|\omega_n)$ given \mathcal{G}_{n+1} . Even if it is, it may not be everywhere proper. Thus, a Kolmogorov conditionalizer may not be able to comply with Fig. 10 even when rcds $P_{\mathcal{G}_1}, P_{\mathcal{G}_2}, \dots, P_{\mathcal{G}_n}, \dots$ exist. Luckily, though, Sokal (1981) has proved that my assumptions are satisfiable in numerous cases, including all or virtually all cases likely to arise in scientific applications. Sokal shows that, under rather mild conditions, there exist $P_{\mathcal{G}_1}, P_{\mathcal{G}_2}, \dots, P_{\mathcal{G}_n}, \dots$ such that, for all n ,

$$\begin{aligned} P_{\mathcal{G}_n} & \text{ is an rcd for } P \text{ given } \mathcal{G}_n \\ P_{\mathcal{G}_n} & \text{ is proper at } \omega, \text{ for all } \omega \in \Omega \\ P_{\mathcal{G}_{n+1}} & \text{ is an rcd for } P_{\mathcal{G}_n}(\cdot|\omega) \text{ given } \mathcal{G}_{n+1}, \text{ for all } \omega \in \Omega \end{aligned}$$

Sokal’s theorem ensures that, in numerous cases, a Kolmogorov conditionalizer can update her credences in accord with Fig. 10.¹⁰

More specifically, consider a standard setup from Bayesian statistics (Florens et al., 1990): we start with a *parameter space* (A, \mathcal{A}) and a *sample space* (S, \mathcal{S}) and form the product space $(A, \mathcal{A}) \otimes (S, \mathcal{S}) =_{df} (A \times S, \mathcal{A} \otimes \mathcal{S})$.¹¹ The parameter space (A, \mathcal{A}) models possible states of the worlds. The sample space (S, \mathcal{S}) models evidence the agent may receive. In many applications, (S, \mathcal{S}) has the form:

$$(T_1, \mathcal{T}_1) \otimes (T_2, \mathcal{T}_2) \otimes (T_3, \mathcal{T}_3) \dots \otimes (T_n, \mathcal{T}_n) \dots$$

and models a stream of incoming evidence received at times $t_1, t_2, \dots, t_n, \dots$. Each outcome then has the form

$$\omega = (\omega_0, \omega_1, \omega_2, \omega_3, \dots, \omega_n, \dots),$$

where $\omega_0 \in A$ and $\omega_n \in T_n$ for $n > 0$. Define random variable $X_n: A \times S \rightarrow T_n$ by projection onto the n th coordinate:

$$X_n(\omega) = \omega_n.$$

¹⁰ Here is the theorem proved by Sokal: Let (Ω, \mathcal{F}, P) be a probability space, where P is perfect and \mathcal{F} is countably generated. Let $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3 \dots, \mathcal{G}_n, \dots$ be a sequence of sub- σ -fields such that $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$ for every n . Suppose that there exist functions Φ_n such that (a) Φ_n is a selection homomorphism for \mathcal{G}_n with respect to \mathcal{F} , for every n ; and (b) $\Phi_n \circ \Phi_{n+m} = \Phi_n$, for every n, m . Then there exist $P_{\mathcal{G}_1}, P_{\mathcal{G}_2}, \dots, P_{\mathcal{G}_n}, \dots$ such that for every n : (i) $P_{\mathcal{G}_n}$ is an rcd for P given \mathcal{G}_n ; (ii) $P_{\mathcal{G}_n}$ is proper at ω , for every ω ; and (iii) $P_{\mathcal{G}_n}(A|\omega) = \int P_{\mathcal{G}_{n+1}}(A|\nu)P_{\mathcal{G}_n}(d\nu|\omega)$, for all $A \in \mathcal{F}$ and all $\omega \in \Omega$. It follows that $P_{\mathcal{G}_{n+1}}$ is an rcd for $P_{\mathcal{G}_n}(\cdot|\omega)$ given \mathcal{G}_{n+1} , for every $\omega \in \Omega$.

¹¹ $\mathcal{A} \otimes \mathcal{S}$ is the σ -field generated by sets of the form $D \times E$, where $D \in \mathcal{A}$ and $E \in \mathcal{S}$.

To model sequential evidence accumulation regarding the sample spaces (T_n, \mathcal{T}_n) , we use certainty profiles over the sub- σ -fields

$$\sigma(X_1), \sigma(X_1, X_2), \sigma(X_1, X_2, X_3), \dots, \sigma(X_1, X_2, \dots, X_n), \dots$$

In this setting, Sokal’s theorem applies whenever the probability space $(A \times S, \mathcal{A} \otimes \mathcal{S}, P)$ meets mild conditions—conditions that are almost always met in the daily practice of Bayesian statistics.¹² Assuming the conditions met, there exist functions

$$P_{\sigma(X_1)}, P_{\sigma(X_1, X_2)}, P_{\sigma(X_1, X_2, X_3)}, \dots, P_{\sigma(X_1, X_2, \dots, X_n)}, \dots$$

such that, for all n ,

$P_{\sigma(X_1, X_2, \dots, X_n)}$ is an rcd for P given $\sigma(X_1, X_2, \dots, X_n)$

$P_{\sigma(X_1, X_2, \dots, X_n)}$ is proper at ω , for all $\omega \in \Omega$

$P_{\sigma(X_1, X_2, \dots, X_{n+1})}$ is an rcd for $P_{\sigma(X_1, X_2, \dots, X_n)}(\cdot|\omega)$ given $\sigma(X_1, X_2, \dots, X_{n+1})$, for all $\omega \in \Omega$

A Kolmogorov conditionalizer with these initial conditional and unconditional credences can update her credences in compliance with Generalized Rigidity.

Here is a simple example of Generalized Rigidity in action.

Rabies Infection Pierre, who lives in nineteenth century Paris, is bit by a rabid dog on his 35th birthday. Knowing that rabies is 100% fatal, Pierre becomes certain that he will die before his 37th birthday. A week later, Pierre learns that Louis Pasteur has invented a vaccine for rabies and that the vaccine is highly effective if delivered soon enough after a bite by a rabid animal. Pierre contacts Pasteur and receives the vaccine. Pierre is not convinced that the vaccine will work, but he is no longer certain that he will die before his 37th birthday.

Using Generalized Rigidity, we can elaborate Rabies Infection into a model that includes principled acquisition and loss of certainties. Let Ω be a suitable outcome space, and let

- Rabies* = the set of outcomes in which Pierre is infected with rabies on his 35th birthday.
- Vaccine* = the set of outcomes in which Pierre receives an effective rabies vaccine within a week after his 35th birthday
- Dead* = the set of outcomes in which Pierre dies before his 37th birthday

¹² The functions Φ_n presupposed by Sokal’s theorem are easily shown to exist in the present context (Sokal, 1981, p. 544). Thus, one need only assume that P is perfect and that $\mathcal{A} \otimes \mathcal{S}$ is countably generated.

Let \mathcal{F} be the σ -field generated by $\{Rabies, Vaccine, Dead\}$. Suppose Pierre has initial conditional and unconditional credences before his 35th birthday given by:

$$\begin{aligned} P(Rabies) &= 1/30,000,000 \\ P(Vaccine) &= 0 \\ P(Dead|Rabies \cap Vaccine) &= 1/2 \\ P(Dead|Rabies \cap Vaccine^c) &= 1 \\ P(Dead|Rabies^c) &= P(Dead|Rabies^c \cap Vaccine) = 1/500 \end{aligned}$$

These credences reflect the following factors: the low rate of exposure to rabies in nineteenth century Paris; non-existence of an effective rabies vaccine, so far as Pierre initially knows; the chancy nature of a new vaccine for a fatal illness; the certain death that follows from a rabies infection absent an effective treatment; and the chance of death from other causes. Assume that $P(A|A) = 1$ for all A . Through the law of total probability and the ratio formula, our assumptions determine a unique probability measure P over \mathcal{F} . For example, we compute:

$$\begin{aligned} P(Dead) &= P(Dead|Rabies \cap Vaccine)P(Rabies \cap Vaccine) \\ &+P(Dead|Rabies \cap Vaccine^c)P(Rabies \cap Vaccine^c) \\ &+P(Dead|Rabies^c)P(Rabies^c) \\ &= 0 + 1/30,000,000 + 1/500 \times 29,999,999/30,000,000 \approx .00200003327 \end{aligned}$$

and

$$P(Dead|Rabies) = P(Dead|Rabies \cap Vaccine^c) = 1.$$

Let \mathcal{G}_1 be the σ -field generated by $\{Rabies\}$, and let \mathcal{G}_2 be the σ -field generated by $\{Vaccine, Rabies\}$. Our stipulations determine a privileged rcd for P given \mathcal{G}_1 , defined by

$$P_{\mathcal{G}_1}(A|\omega) = \begin{cases} P(A|Rabies) & \text{if } \omega \in Rabies \\ P(A|Rabies^c) & \text{if } \omega \notin Rabies \end{cases}$$

and a privileged rcd for P given \mathcal{G}_2 , defined by

$$P_{\mathcal{G}_2}(A|\omega) = \begin{cases} P(A|Rabies \cap Vaccine) & \text{if } \omega \in Rabies \cap Vaccine \\ P(A|Rabies \cap Vaccine^c) & \text{if } \omega \in Rabies \cap Vaccine^c \\ P(A|Rabies^c \cap Vaccine) & \text{if } \omega \in Rabies^c \cap Vaccine \\ P(A|Rabies^c \cap Vaccine^c) & \text{if } \omega \in Rabies^c \cap Vaccine^c \end{cases}$$

It is not hard to check that, for every ω , $P_{\mathcal{G}_2}$ is an rcd for $P_{\mathcal{G}_1}(\cdot|\omega)$ given \mathcal{G}_2 . Let ω be the true outcome. When the rabid dog bites Pierre, he becomes certain of *Rabies* and responds by forming new credences $P_{\mathcal{G}_1}(\cdot|\omega)$. In particular, he becomes certain of *Dead*. Complying with Generalized Rigidity, his credences conditional on \mathcal{G}_2 are

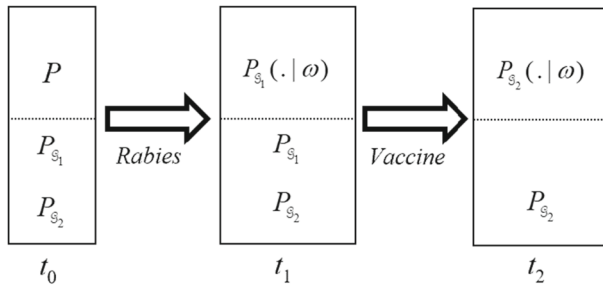


Fig. 11 Credal evolution in Rabies Infection. ω is the true outcome

still given by $P_{\mathcal{G}_2}$. When he receives the rabies vaccine, he becomes certain of *Vaccine* and responds by forming new credences $P_{\mathcal{G}_2}(\cdot | \omega)$, so that his credence in *Dead* goes from 1 to $\frac{1}{2}$. See Fig. 11.

Rabies Infection illustrates the advantages that Weak Generalized Rigidity offers over Weak Rigidity. Pierre’s initial conditional credences carry forward in accord with Weak Generalized Rigidity. The conditional credences determine how he should conditionalize both when he learns that he is infected with rabies *and* when he later learns that he has received a rabies vaccine. In this manner, Weak Generalized Rigidity helps us model situations where the agent accumulates evidential certainties (modeled by certainties over \mathcal{G}_1 and then over \mathcal{G}_2). Weak Rigidity only helps us model situations where evidence gained at a later time eradicates evidential certainties gained at an earlier time (modeled by shifting certainties over \mathcal{G}_1).

Relatedly, Weak Generalized Rigidity is much more useful than Weak Rigidity when the factivity assumption (3) prevails. To apply Weak Rigidity in a non-trivial way, we must consider a scenario where a certainty profile δ_v over \mathcal{G} supplants a conflicting certainty profile δ_ω over \mathcal{G} . Conflicting certainty profiles cannot both satisfy the factivity assumption. At least one of them must be misplaced. In contrast, Generalized Rigidity helps us model cases where the agent conditionalizes on a certainty profile that satisfies the factivity assumption and subsequently conditionalizes on a distinct certainty profile that also satisfies the factivity assumption. Pierre correctly becomes certain that he is infected with rabies, then correctly becomes certain that he received an effective rabies vaccine. At neither time does he acquire misplaced certainties over a conditioning sub- σ -field. Nevertheless, evidence gained at the later time defeats evidence gained at the earlier time. He gains strong evidence that he will die before his 37th birthday (he was infected by rabies), then subsequently receives strong defeating evidence (he has received an effective rabies treatment). Thus, Generalized Rigidity helps us model cases of epistemic defeat where the defeated evidence is veridical and the defeating evidence is also veridical.¹³

¹³ Some readers may protest that, while my model of Rabies Infection does not explicitly invoke non-factive conditionalization, Pierre’s initial certainty that no effective rabies vaccine exists must result somehow from non-factive conditionalization. I disagree. Pierre may have come to set $P(\text{Vaccine}) = 0$ based solely on veridical evidence. For example, let *Expert* be the proposition that the renowned medical expert Dr. Charbonnet asserts that no effective rabies vaccine could ever exist. Suppose Pierre has complete faith in Charbonnet and therefore sets $P(\text{Vaccine}|\text{Expert}) = 0$. If Charbonnet tells Pierre that no effective rabies vaccine could ever exist, Pierre will conditionalize on the true proposition *Expert* and hence set $P(\text{Vaccine})$

Here is a more elaborate example along similar lines.

Ring Time A timer will ring during an interval $[x, x + y]$. Frank chooses the start time x , and Mary independently chooses the interval length y . Sarah believes that, given x and y , the timer is equally likely to ring at any time z falling in the interval $[x, x + y]$. At t_0 , Sarah is certain that Frank will choose a start time between 0 and 100, and she believes that he is equally likely to choose any start time within that interval. She is certain that Mary will choose interval length y_0 . At time t_1 , Sarah learns that Frank chose start time $x_1 \in [0, 100]$. This discovery, combined with her certainty that the interval length is y_0 , leads Sarah to become certain that the timer will ring during the interval $[x_1, x_1 + y_0]$. At t_2 , Sarah learns that Mary chose interval length $y_2 > y_0$. This discovery eradicates Sarah’s certainty that the timer will ring during the interval $[x_1, x_1 + y_0]$.

Using Generalized Rigidity, we can fill in the story to rationalize Sarah’s certainty loss at t_2 . We use a probability space $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2), \pi)$ to codify Sarah’s credences over possible start times x and interval lengths y . We use $C: \mathcal{B}(\mathbb{R}) \times \mathbb{R}^2 \rightarrow \mathbb{R}$ to codify Sarah’s credences over ring times *conditional on start time x and interval length y* . Given how I described Sarah’s initial credences, we naturally choose π and C defined by

$$\pi =_{df} U_{[0, 100]} \otimes \delta_{y_0}$$

$$C(\cdot | x, y) =_{df} U_{[x, x+y]}$$

where \otimes is the product measure (Billingsley, 1995, pp. 232–233) and $U_{[a, b]}$ is the uniform distribution over $[a, b]$. We define a probability measure P over the larger space $(\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3))$:

$$P(A) = \int \left[\int I_A(x, y, z) C_{x,y}(dz) \right] d\pi(x, y), \quad \text{for any } A \in \mathcal{B}(\mathbb{R}^3) \tag{17}$$

where I_A is the indicator function for A :

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

P encodes Sarah’s credences over start times x , interval lengths y , and ring times z .¹⁴ It is easy to show that, for our choice of C and π , the measure P defined by (17) concentrates all probability mass over the event

Footnote 13 continued

= 0. Thus, the credences posited by my model may arise through the exercise of factive conditionalization. Whether Pierre is *rationally permitted* to set $P(\text{Vaccine} | \text{Expert}) = 0$ is a trickier question that I must set aside for limitations of space.

¹⁴ Definition (17) is a special case of the *Ionescu Tulcea extension theorem*, which shows that under highly general conditions one can convert conditional probabilities and select unconditional probabilities into a global unconditional probability measure (Kallenberg, 2002, p. 116).

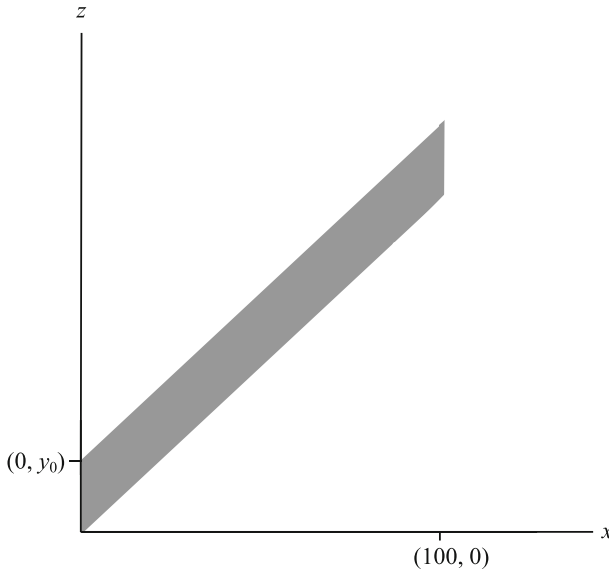


Fig. 12 Sarah’s credences at t_0 . The whole probability space is three-dimensional, but Sarah’s initial credal mass lies entirely in the two-dimensional sub-space (pictured here) where $Y = y_0$. Sarah assigns credence 1 to the grey parallelogram

$$\{(x, y_0, z) : 0 \leq x \leq 100 \ \& \ x \leq z \leq x + y_0\}.$$

See Fig. 12. Let X and Y be projection mappings onto the x and y coordinates, respectively. Then the following function $P_{\sigma(X, Y)} : \mathcal{B}(\mathbb{R}^3) \times \mathbb{R}^3 \rightarrow \mathbb{R}$ is an rcd for P given $\sigma(X, Y)$:

$$P_{\sigma(X, Y)}(\cdot | x, y, z) = \delta_x \otimes \delta_y \otimes U_{[x, x+y]}.$$

Note that $P_{\sigma(X, Y)}$ embeds the conditional credences C into corresponding conditional credences over the larger space $(\mathbb{R}^3, \mathcal{B}(\mathbb{R}^3))$. One can also show that the following function $P_{\sigma(X)} : \mathcal{B}(\mathbb{R}^3) \times \mathbb{R}^3 \rightarrow \mathbb{R}$ is an rcd for P given $\sigma(X)$:

$$P_{\sigma(X)}(\cdot | x, y, z) = \delta_x \otimes \delta_{y_0} \otimes U_{[x, x+y_0]}.$$

Intuitively: newfound certainty that the start time is x , combined with prior certainty that $Y = y_0$, induces a uniform distribution over ring times falling in the interval $[x, x + y_0]$. In addition, one can show that $P_{\sigma(X, Y)}$ is an rcd for $P_{\sigma(X)}(\cdot | x, y, z)$ given $\sigma(X, Y)$, for any x, y, z .¹⁵

At t_0 , Sarah has unconditional credences codified by P along with conditional credences codified by $P_{\sigma(X)}$ and $P_{\sigma(X, Y)}$. At t_1 , she becomes certain that Frank chose

¹⁵ Proofs for all mathematical claims made in this paragraph are straightforward but tedious.

start time $X = x_1$. We model her newfound certainty through a certainty profile $\delta_{(x_1, y, z)}$ over $\sigma(X)$, where y and z are arbitrary. We may assume that Frank really did choose x_1 , so that $\delta_{(x_1, y, z)}$ satisfies the factivity assumption. Sarah conditionalizes using $P_{\sigma(X)}$, forming new credences:

$$P_{\sigma(X)}(\cdot | x_1, y, z) = \delta_{x_1} \otimes \delta_{y_0} \otimes U_{[x_1, x_1 + y_0]}.$$

Thus, she is newly certain of the event

$$\{x_1\} \times \{y_0\} \times [x_1, x_1 + y_0].$$

In other words: she is certain that John chose $X = x_1$, that Mary chose $Y = y_0$, and that the timer will ring in the interval $[x_1, x_1 + y_0]$. See Fig. 13. Complying with Generalized Rigidity, she retains $P_{\sigma(X)}$ and $P_{\sigma(X, Y)}$ as conditional credences at t_1 . At t_2 , she becomes certain that Mary chose the interval length $y_2 > y_0$. We model Sarah’s newfound certainty through a certainty profile $\delta_{(x_1, y_2, z)}$ over $\sigma(X, Y)$. We may assume that Sarah is correct, so that $\delta_{(x_1, y_2, z)}$ satisfies the factivity assumption. In response to her new certainty profile, she conditionalizes using $P_{\sigma(X, Y)}$. Her credences are now given by:

$$P_{\sigma(X, Y)}(\cdot | x_1, y_2, z) = \delta_{x_1} \otimes \delta_{y_2} \otimes U_{[x_1, x_1 + y_2]}.$$

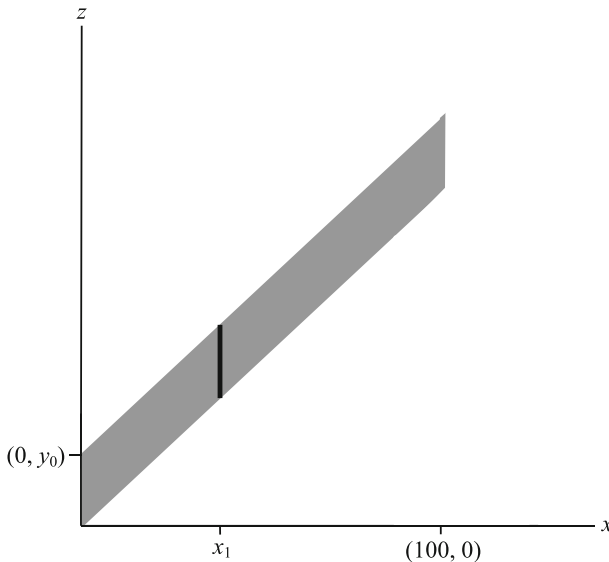


Fig. 13 Sarah’s credences at t_1 . She assigns credence 1 to the event (pictured here by a black line segment) formed by intersecting the line $X = x_1$ with the grey parallelogram

See Fig. 14. Thus, Sarah is newly certain of the event

$$\{x_1\} \times \{y_2\} \times [x_1, x_1 + y_2].$$

See Fig. 15. Her newfound certainty that $Y = y_2$ eradicates her certainty (gained at t_1) that the timer will ring in the interval $[x_1, x_1 + y_0]$.

In both Rabies Infection and Ring Time, the agent’s credences evolve according to Fig. 9. Some certainties gained at t_1 are subsequently lost at t_2 . The examples demonstrate that Kolmogorov conditionalization, aided by Weak Generalized Rigidity, can induce principled acquisition and loss of certainties even when the factivity assumption (3) prevails.

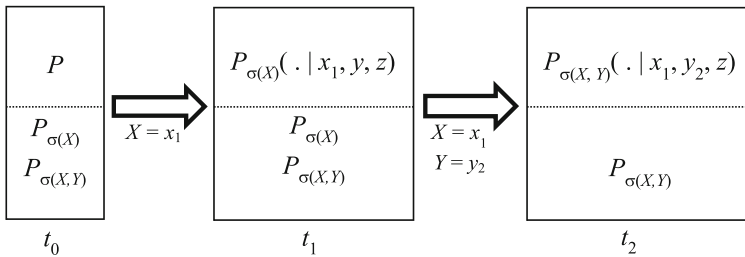


Fig. 14 Credal evolution in Ring Time

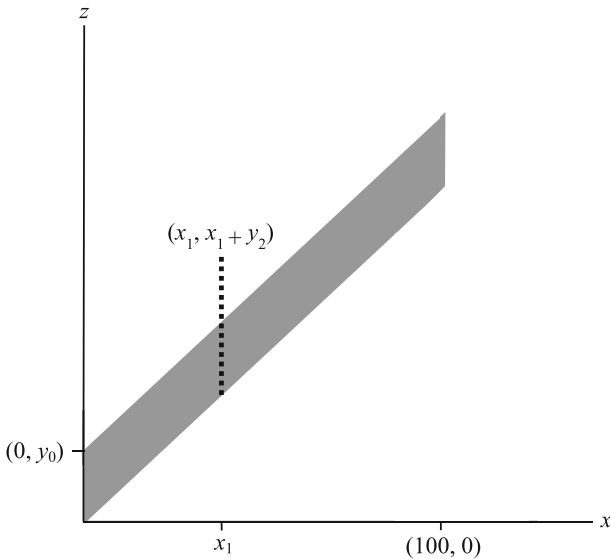


Fig. 15 Sarah’s credences at t_2 . Her credal mass lies entirely in the two-dimensional sub-space (pictured here) where $Y = y_2$. She assigns credence 1 to the dotted black line segment reaching to $(x_1, x_1 + y_2)$

8 Scope and limits

I have illustrated Weak Rigidity and Weak Generalized Rigidity with four toy examples: False Alarm, Mismeasurement, Rabies Infection, and Ring Time. The toy examples only hint at the scope of Fig. 10. Sokal's theorem guarantees the existence of suitable rcds across diverse circumstances, including the vast majority of cases likely to arise in scientific applications. Whenever suitable rcds exist, a Kolmogorov conditionalizer can sequentially update her credences while complying with Weak Generalized Rigidity. Quite often, the sequential credal updates will eradicate certainties gained through earlier credal updates.

Although Weak Generalized Rigidity is widely applicable, there are situations where it does not apply. Weak Generalized Rigidity imposes a substantive constraint only when the agent gains new certainties over a conditioning sub- σ -field \mathcal{G} . As Jeffrey emphasizes, though, an agent may gain new credences over \mathcal{G} that are not certainties. Jeffrey focuses on situations where \mathcal{G} is generated by a countable partition. The general case, where \mathcal{G} is not necessarily generated by a countable partition, has received some attention (e.g. Diaconis & Zabell, 1982; Hild et al., 1999; Meehan & Zhang, 2020) but not as much as it deserves. I suspect that Generalized Rigidity can shed light upon the general case. In any event, there are plainly situations where Weak Generalized Rigidity offers little help.

Even if we restrict attention to credal changes sparked by new certainties, Weak Generalized Rigidity does not always offer useful guidance. Consider a variant of Mismeasurement: Jane learns at t_2 that her laboratory assistant (who reported the result of the experiment measuring X 's value) is a pathological liar. Clearly, Jane should downgrade her certainty in $X = x_1$. Assuming that X does in fact have value x_1 , we cannot model this case using Weak Generalized Rigidity. Weak Generalized Rigidity imposes a substantive constraint on cases falling into two categories:

- (i) An agent acquires a certainty profile over a sub- σ -field \mathcal{G} , then subsequently acquires a different certainty profile over \mathcal{G} , and so on.
- (ii) An agent acquires a certainty profile over a sub- σ -field \mathcal{G}_1 , then acquires a certainty profile over a sub- σ -field \mathcal{G}_2 such that $\mathcal{G}_1 \subseteq \mathcal{G}_2$, and so on.

Our new variant of Mismeasurement does not fall under either (i) or (ii):

- (i) Distinct certainty profiles over \mathcal{G} cannot both conform to the factivity assumption. So category (i) does not include cases where the agent gains true evidence at every stage.
- (ii) If certainty profiles over \mathcal{G}_1 and \mathcal{G}_2 both satisfy the factivity assumption, and $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then certainties over \mathcal{G}_1 gained at t_1 persist when the agent gains new certainties over \mathcal{G}_2 at t_2 .

Neither category (i) nor category (ii) includes cases where true evidence eradicates certainty in a true conditioning proposition. So my framework cannot model how Jane's credences change in response to learning that her lab assistant is a pathological liar. More generally, my framework does not help us model an agent who conditionalizes on a true proposition E but later learns truths that defeat her warrant for E .¹⁶

¹⁶ See (Weisberg, 2009a) for related discussion in connection with ratio conditionalization and Jeffrey conditionalization.

In response, one option is to abandon the assumption that the sub- σ -fields are non-decreasing. Rather than consider a sequence of sub- σ -fields such that

$$\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{G}_3 \dots \subseteq \mathcal{G}_n \subseteq \dots,$$

we can instead consider a sequence of sub- σ -fields

$$\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3 \dots, \mathcal{G}_n, \dots$$

that is not necessarily non-decreasing. An agent may conditionalize based on new certainties over \mathcal{G}_1 , then conditionalize based on new certainties over \mathcal{G}_2 , and so on. Kallenberg (2010) has investigated iterated Kolmogorov conditionalization in this alternative setting. The alternative setting does not assume that $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$, so Generalized Rigidity does not apply. In the alternative setting, the agent can gain certainties over \mathcal{G}_n and subsequently lose those certainties in response to new certainties over \mathcal{G}_{n+1} *even though all certainties conform to the factivity assumption*. In future work, I will use the alternative setting to model cases where true evidence eradicates newfound certainty in a true conditioning proposition.

Clearly, Fig. 10 is not general enough to handle *all* cases of epistemic defeat. Still, it is general enough to handle *many* cases of epistemic defeat. In that respect, it marks significant progress over theories that rely solely on ratio conditionalization.

9 Comparisons

To clarify the scope and limits of my approach, I will now compare it with three treatments found in the literature. The treatments are due respectively to Skyrms (1983), Titelbaum (2013), and Williamson (2000).

9.1 Skyrms on memory

Skyrms (1983, p. 157) notes that, when we update credences through ratio conditionalization, “there is a certain peculiar sense in which we lose information every time we learn something. That is, we lose information concerning the initial relative probabilities of statements not entailing S ” whenever we ratio conditionalize on S . He proposes that we “give a probability assignment a memory” (p. 157), so as to retain information that would otherwise be lost. He suggests two information retention strategies. The basic idea behind both strategies is to maintain a record of earlier probabilities P , so that conditional probabilities $P(H|F)$ can be computed using the ratio formula even after conditionalizing on a proposition E incompatible with F . The first strategy records unconditional probabilities from each time stage. The second strategy records initial unconditional probabilities along with total evidence to date.

By maintaining a record of previous unconditional probabilities and using that record to compute conditional probabilities, Skyrms’s treatment reflects roughly the same “minimal change ethos” as Rigidity and Generalized Rigidity. However, his proposed strategies retain more information than is needed to handle the defeasible

inferences analyzed in this paper. If an agent conditionalizes on E and subsequently wishes to conditionalize on an incompatible F , then she does not need access to her initial unconditional probabilities. She only needs access to suitable probabilities conditional on F . By retaining earlier unconditional probability assignments, Skyrms's two strategies retain extraneous information. In contrast, my approach retains only the needed conditional probabilities. In Mismeasurement, for example, Jane can conditionalize on $X = x_1$ at t_1 and then (in accord with Weak Rigidity) use the same conditional probabilities to conditionalize on $X = x_2$ at t_2 . Jane's unconditional probabilities from t_0 do not matter at t_2 . All that matters are her t_1 conditional probabilities, as enshrined by $P_{\sigma(X)}$.

More importantly, Skyrms's two strategies do not handle examples featuring initial conditional probabilities beyond the ratio formula. We have seen that such examples arise routinely in scientific practice. We have also seen that the standard mathematical and scientific solution is to use rcds. My proposal builds upon the standard solution, carrying forward conditional probabilities as codified by rcds in accord with Weak Generalized Rigidity. In this way, my approach handles numerous cases that Skyrms's does not, such as Rabies Infection, Ring Time, and versions of Mismeasurement where X has uncountably many possible values.¹⁷

Skyrms writes that, “[a]fter conditionalizing on S , one might wish to be able to decide that this was an error and ‘deconditionalize’” (1983, p. 157). He models deconditionalization using the second of his information retention strategies: the agent adds proposition S to her total evidence at one time stage and deletes S from her total evidence at a later time stage (p. 159); conditionalizing on total evidence at the later time stage yields the desired deconditionalization. Notably, though, Skyrms does not formally model the factors that impel the agent to delete S from her total evidence. For that reason, his proposal does not model everything we would like to model about deconditionalization.

The situation is roughly comparable if we employ the rcd formalism. Consider yet another variant of Mismeasurement, in which Jane decides at t_2 that it was a mistake to conditionalize on $X = x_1$ and retreats to her former uncertainty regard X 's value. By stipulation, Jane's t_2 credences over $\sigma(X)$ are given by P . In accord with Rigidity, she carries forward her conditional probabilities $P_{\sigma(X)}$ from t_1 to t_2 . Then it is straightforward to show that her t_2 credences over the entire space are given by P . Thus, Jane's renewed uncertainty regarding X 's value carries her back to her t_0 credences, as one would intuitively expect. In this manner, the rcd formalism conjoined with Rigidity helps us model Jane's evolution from uncertainty regarding

¹⁷ Skyrms mentions a third strategy that uses Jeffrey conditionalization rather than ordinary conditionalization. This third strategy may be useful for some purposes, but I do not think it fits current scientific practice very well. As indicated in Sect. 1, scientific applications of the Bayesian framework almost exclusively feature inferences that employ ordinary conditionalization rather than Jeffrey conditionalization. Virtually all of those inferences are vulnerable to epistemic defeat (e.g. a scientist may update based on the wrong value of an experimental variable, as in Mismeasurement). Replacing ordinary conditionalization with Jeffrey conditionalization would leave us unable to model such inferences as they occur in current practice. In contrast, my account preserves the central role that ordinary conditionalization plays within current scientific practice while accommodating types of epistemic defeat not typically explicitly addressed by that practice. My account *extends* current scientific practice, whereas an account grounded in Jeffrey conditionalization *revises* current scientific practice.

X 's value to certainty and then back to uncertainty. But the result is arguably not too satisfying because it does not explicitly model *why* Jane decides that it was a mistake to conditionalize on $X = x_1$. A more satisfying treatment would explicitly model Jane's t_2 evidence and how that evidence reinstates her t_0 credences over $\sigma(X)$. I believe that, ultimately, deconditionalization requires us to grapple with the issues raised in Sect. 8.

9.2 Titelbaum on certainty loss

Titelbaum (2013) models certainty eradication within a broadly Bayesian framework. He advances a diachronic norm, *Generalized Conditionalization*, that generalizes ratio conditionalization but can also accommodate some cases where agents lose certainties. He admits, though, that his framework does not satisfactorily handle "cases in which an agent becomes certain of a claim in response to a piece of evidence, then withdraws that certainty upon encountering a defeater" (pp. 296–298).

Suppose that evidence e gained at t_1 leads an agent to become certain of hypothesis h and that further evidence d gained at t_2 eradicates her newfound certainty in h . Titelbaum's framework allows this to happen, but only within a model that *stipulates* certainty in h at t_1 . Such a model cannot capture why the agent's doxastic state at t_0 mandates her certainty in h at t_1 in response to evidence e . As Titelbaum puts it, "nothing in the model is tracking e 's influence on h " (p. 297). In contrast, my framework can model how an agent's initial credences (both conditional and unconditional) mandate new certainties at t_1 and loss of those certainties at t_2 . False Alarm, Mismeasurement, Rabies Infection, and Ring Time all have that form. In each case, my model tracks e 's influence on h at t_1 (e.g. learning that he was bit by a rabid dog leads Pierre to become certain that he will soon die of rabies) and d 's influence on h at t_2 (e.g. learning that he received a rabies vaccine leads Pierre to lose that newfound certainty). So my framework supplies substantive rational constraints beyond those supplied by Titelbaum's.

9.3 Williamson on evidence

Williamson (2000, pp. 205–206) suggests that you can gain evidence and assign it probability 1, then gain new evidence *that has non-zero probability* and on that basis rationally downgrade the probability of the old evidence. His main example runs as follows. You inspect a red ball and a black ball before placing them in an empty bag, which leads you to become certain that you placed a red ball and a black ball in the bag. You then execute 10,000 draws with replacement that all turn out red. According to Williamson, you should lose your certainty that you placed a black ball in the bag.

The proposition that you drew red 10,000 times has non-zero probability, so conditionalizing on that proposition cannot dislodge your certainty that you placed a black ball in the bag. Thus, as Williamson (pp. 219–220) emphasizes, his analysis mandates major revisions to Bayesian decision theory. He proposes an alternative framework that retains some Bayesian elements while rejecting the core Bayesian idea that agents

should respond to new evidence by conditionalizing on it (p. 220). Williamson's framework is designed to handle the bag example and other examples where evidence with non-zero probability supposedly dislodges certainties.

I see no reason to revise Bayesian decision theory along the lines suggested by Williamson. On the contrary, the standard Bayesian analysis of the bag example seems correct to me: if you really are certain that you placed a black ball in the bag, then drawing red 10,000 times cannot rationally dislodge that certainty. Of course, it is very unlikely that you would draw red 10,000 times if you placed a black ball in the bag. But unlikely events sometimes occur, and they are not in themselves a sound basis for certainty eradication. Perhaps you should not have initially become certain that you placed a black ball in the bag, yet *given that you did become certain* your certainty should persist despite the 10,000 red draws. Here it is crucial to distinguish between certainty *that you placed a black ball in the bag* and certainty *that the bag currently contains a black ball*. Even if you are certain of the first proposition, it does not follow that you are (or should be) certain of the second. After all, the black ball might have surreptitiously escaped through a small hole or some other chicanery. Drawing red 10,000 times may rationally lead you to downgrade your credence that the bag *currently* contains a black ball while remaining certain that you placed a black ball in the bag.

At any rate, my framework cannot model situations (if such there are) where evidence with non-zero probability rationally dislodges certainties. It *can* model numerous situations where evidence with probability zero rationally dislodges certainties. It models these situations without any revision to Bayesian decision theory. On my approach, certainties cannot be dislodged by mere *unlikely* evidence, but they can be dislodged by *vanishingly unlikely* evidence.

10 Conclusion

Philosophers should retire the canard that conditionalization always leaves certainties in place. This canard reflects an overly narrow focus on the ratio formula and distorts scientific practice. Kolmogorov conditionalization can eradicate certainties, including certainties gained through prior exercise of conditionalization. It generates reasonable credal updates across a range of cases, especially when supplemented with Weak Generalized Rigidity. Overall, then, Kolmogorov conditionalization offers significant advantages over ratio conditionalization, including an improved ability to model epistemic defeat.

I do not say that Kolmogorov conditionalization provides a universal basis for defeasible inference. No doubt there are situations where Jeffrey conditionalization or some other credal update strategy would be more appropriate. There are also situations where Kolmogorov conditionalization is inapplicable, such as when reds do not exist or are improper. Still, Kolmogorov conditionalization is a valuable addition to the Bayesian toolbox. Further work will surely reveal additional philosophical applications beyond those discussed here. I submit that reds have great potential to enrich formal epistemology, as they have already enriched other disciplines where Bayesian decision

theory plays a prominent role. There awaits a vast terrain that philosophers have barely explored.

Acknowledgements I am indebted to Kenny Easwaran, Simon Huttegger, Alexander Meehan, Sean Walsh, Snow Zhang, and two anonymous referees for this journal for helpful discussion and feedback. I presented some of this material in 2018 at a symposium session of the 26th Biennial Meeting of the Philosophy of Science Association. I thank all those who participated, especially Miklós Rédei and Brian Skyrms.

Declarations

Conflict of interest There are no conflicts of interest, disclosures, or funding sources to declare.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bennett, B., Hoffman, D., Prakash, C., & Richman, S. (1996). Observer theory, Bayes theory, and psychophysics. In D. Knill & W. Richards (Eds.), *Perception as Bayesian inference*. Cambridge University Press.
- Billingsley, P. (1995). *Probability and Measure* (3rd ed.). Wiley.
- Blackwell, D., & Dubins, L. (1975). On existence and non-existence of proper, regular conditional distributions. *The Annals of Probability*, 3, 741–752.
- Blackwell, D., & Ryll-Nardzewski, C. (1963). Non-existence of everywhere proper conditional distributions. *The Annals of Mathematical Statistics*, 34, 223–225.
- Diaconis, P., & Zabell, S. (1982). Updating subjective probability. *The Journal of the American Statistical Association*, 77, 822–830.
- Doucet, A., de Freitas, N., & Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer.
- Dubins, L. (1975). Finitely additive conditional probabilities, conglomerability and disintegrations. *Annals of Probability*, 3, 89–99.
- Earman, J. (1992). *Bayes or bust?* MIT Press.
- Easwaran, K. (2008). *The Foundations of Conditional Probability*. PhD dissertation, University of California, Berkeley. Ann Arbor: ProQuest/UMI. (Publication No. 3331592.).
- Easwaran, K. (2011). Varieties of conditional probability. In P. Bandyopadhyay and M. Forster (eds) *Philosophy of Statistics*, Burlington: Elsevier.
- Easwaran, K. (2019). Conditional probabilities. In R. Pettigrew & J. Weisberg (Eds.), *The Open handbook of formal epistemology*. PhilPapers.
- Feldman, M. (1987). Bayesian learning and convergence to rational expectations. *Journal of Mathematical Economics*, 16, 292–313.
- Florens, J.-P., Mouchart, M., & Rolin, J.-M. (1990). *Elements of Bayesian statistics*. Marcel Dekker Inc.
- Ghosal, S., & van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*. Cambridge University Press.
- Griffiths, T., Kemp, C., & Tenenbaum, J. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology*. Cambridge University Press.
- Gyenis, Z., & Rédei, M. (2017). General properties of Bayesian learning as statistical inference determined by conditional expectations. *The Review of Symbolic Logic*, 10, 719–755.
- Hild, M., Jeffrey, R., & Risse, M. (1999). Aumann's "No Agreement" theorem generalized. In C. Bicchieri, R. Jeffrey, & B. Skyrms (Eds.), *The logic of strategy*. Oxford University Press.

- Huttegger, S. (2015). Merging of opinions and probability kinematics. *The Review of Symbolic Logic*, 8, 611–648.
- Huttegger, S., & Nielsen, M. (2020). Generalized learning and conditional expectation. *Philosophy of Science*, 87, 868–883.
- Jeffrey, R. (1983). *The logic of decision* (2nd ed.). University of Chicago Press.
- Joyce, J. (2009). The development of subjective Bayesianism. In D. Gabbay, S. Hartman, & J. Woods (Eds.), *Handbook of the history of logic*. (Vol. 10). Elsevier.
- Kallenberg, O. (2002). *Foundations of modern probability* (2nd ed.). Springer.
- Kallenberg, O. (2010). Commutativity properties of conditional distributions and palm measures. *Communications on Stochastic Analysis*, 4, 21–34.
- Knill, D., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge University Press.
- Kolmogorov, A. N. (1933/1956). *Foundations of the theory of probability*. 2nd English edn. Trans. N. Morrison. New York: Chelsea.
- Levi, I. (1980). *The enterprise of knowledge*. MIT Press.
- Meehan, A., & Zhang, S. (2020). Jeffrey meets Kolmogorov: A general theory of conditioning. *Journal of Philosophical Logic*, 49, 941–979.
- Meehan, A., & Zhang, S. (2022). Kolmogorov conditionalizers can be Dutch booked (if and only if they are evidentially uncertain). *The Review of Symbolic Logic*, 15, 722–757.
- Nielsen, M. (2021). A new argument for Kolmogorov conditionalization. *The Review of Symbolic Logic*, 14, 930–945.
- Popper, K. (1959). *The logic of scientific discovery*. Hutchinson.
- Pryor, J. (2013). Problems for credulism. In C. Tucker (Ed.), *Seemings and justification: new essays on dogmatism and phenomenal conservatism*. Oxford University Press.
- Rao, M. M. (2005). *Conditional measures and applications* (2nd ed.). CRC Press.
- Rényi, A. (1955). On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiarum Hungarica*, 6, 285–335.
- Rescorla, M. (2015a). Bayesian perceptual psychology. In M. Matthen (Ed.), *The Oxford handbook of the philosophy of perception*. Oxford University Press.
- Rescorla, M. (2015b). Some epistemological ramifications of the Borel-Kolmogorov paradox. *Synthese*, 192, 735–767.
- Rescorla, M. (2016). Bayesian sensorimotor psychology. *Mind and Language*, 31, 3–36.
- Rescorla, M. (2018a). A Dutch book theorem and converse Dutch book theorem for Kolmogorov Conditionalization. *The Review of Symbolic Logic*, 11, 705–735.
- Rescorla, M. (2018b). An interventionist approach to psychological explanation. *Synthese*, 195, 1909–1940.
- Rescorla, M. (2020). A realist perspective on Bayesian cognitive science. In A. Nes & T. Chan (Eds.), *Inference and consciousness*. Routledge.
- Rescorla, M. (2021). On the proper formulation of Conditionalization. *Synthese*, 198, 1935–1965.
- Rescorla, M. (2022). An improved Dutch book theorem for Conditionalization. *Erkenntnis*, 87, 1013–1041.
- Rescorla, M. Forthcoming. Non-factive Kolmogorov conditionalization. *The Review of Symbolic Logic*.
- Sanborn, A. (2017). Types of approximation for probabilistic cognition: Sampling and variational. *Brain and Cognition*, 112, 98–101.
- Schervish, M. (1995). *Theory of statistics*. Springer.
- Seidenfeld, T., Schervish, M., & Kadane, J. (2001). Improper regular conditional distributions. *Annals of Probability*, 29, 1612–1624.
- Skyrms, B. (1983). Three ways to give a probability assignment a memory. In J. Earman (Ed.), *Testing scientific theories*. Minneapolis: University of Minnesota Press.
- Sokal, A. (1981). Existence of compatible families of proper regular conditional probabilities. *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete*, 56, 537–548.
- Titelbaum, M. (2013). *Quitting certainties*. Oxford University Press.
- Weisberg, J. (2009a). Commutativity or holism? A dilemma for conditionalizers. *The British Journal for the Philosophy of Science*, 60, 793–812.
- Weisberg, J. (2009b). Varieties of Bayesianism. In D. Gabbay, S. Hartman, & J. Woods (Eds.), *Handbook of the history of logic*. Elsevier.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.