Believing the Unbelievable

John Earman
Dept. HPS
University of Pittsburgh

Bayesian personalism models learning from experience as the updating of an agent's credence function on the information the agent acquires. The standard updating rules are hamstrung for zero probability events. The maneuvers that have been proposed to handle this problem are examined and found wanting: they offer only temporary relief but no satisfying and stable long term resolution. They do suggest a strategy for avoiding the problem altogether, but the price to be paid is a very crabbed account of learning from experience. I outline what Bayesians would need to do in order to come to grips with the problem rather than seeking to avoid it. Furthermore, I emphasize that an adequate treatment of the issues must work not only for classical probability but also for quantum probability as well, the latter of which is rarely discussed in the philosophical literature in the same breath with the updating problem. Since it is not obvious how the maneuvers applied to updating classical probability can be made to work for updating quantum probability a rethinking of the problem may be required. At the same time I indicate that in some special cases quantum probability theory has a self-contained solution to the problem of updating on zero probability events requiring no additional technical devices or rationality constraints.

1 Introduction

It is well known that Bayesianism has a problem with modeling belief revision when an agent is faced with zero probability events. Alleged solutions are gestured at but are rarely critically examined. While it is a bit of an exaggeration to say that the attempted solutions fail utterly, I will argue that the technical machinery thrown at the problem fails to deliver anything approaching a satisfying account of how Bayesian belief revision is supposed to take place in the problem cases. Additionally, while the technical machinery is well developed for classical probability there is, to my knowledge, no discussion of how the machinery can be made to work for quantum probability, which is a serious lacuna since, presumably, the event structure of our world is quantum rather than classical.

I begin by outlining the accepted updating rules for classical and for quantum probability. These rules are stymied in the case of zero probability events. I then turn to a critical examination of the most widely cited rescue efforts and conclude that none of them pass muster. From this discussion emerges a strategy for avoiding the problem altogether, but implementing it comes at the expense of a very crabbed account of learning from experience. A worthy Bayesianism would confront rather than seek to avoid the problem, and this involves facing up to aspects of belief revision which Bayesians have tended to ignore. I outline what needs to be done and offer some options but (wisely? cowardly?) decline to recommend one. Throughout I flag what needs to be done in order to properly pose the issues when credences are assigned in the non-commutative event structure encountered in quantum mechanics (QM).

2 Simple Bayesianism for classical probability

Here I adopt the spirit of Bayesianism personalism: Rational agents have credences in the form of degrees of belief that conform to the axioms of probability with some suitable form of additivity to be discussed; these agents have learning experiences, and in reaction they update their credence functions.[1] The simple form of Bayesianism assumes that what is learned can be encapsulated in the form of a proposition that is learned with certainty. Of course, this is an idealization, and a realistic modeling of a learning experience will have to allow for uncertain learning. Various updating rules have been proposed for uncertain learning, such as Jeffrey conditionalization (see Jeffrey 1965). I will eventually take up Jeffrey conditionalization in Section 8, but initially I will focus on the simple, idealized case since there are already complications aplenty. In the present section I set the issues in the context of classical probability; in the following sections I sketch how the issues at hand play out in the context of quantum probability.

[1] For an overview of the many varieties of Bayesianism see Weisberg (2011).

2.1 Classical probability

A classical probability space is a triple (Ω, Σ, pr) where Ω is the sample space and Σ, the set of measurable sets (i.e. the subsets of Σ to which probability is assigned), is an algebra of subsets of Ω that contains Ω and is closed under complementation and finite unions and intersections. The probability measure pr is a map $\Sigma \to [0,1]$ satisfying at a minimum

(Ci) $pr(\Omega) = 1$

(Cii) $pr(E_1 \cup E_2) = pr(E_1) + pr(E_2)$ for all $E_1, E_2 \in \Sigma$ such that $E_1 \cap E_2 = \emptyset$.

The additivity axiom can be strengthened to

(Cii′) $pr(\cup_{a \in \mathcal{I}} E_a) = \sum_{a \in \mathcal{I}} pr(E_a)$ for all $E_a \in \Sigma$ such that $E_a \cap E_{a'} = \emptyset$ when $a \neq a'$.

When the index set $\mathcal{I}$ is denumerable (Cii′) is the requirement of countable additivity, and when $\mathcal{I}$ can be any cardinality it is the requirement of complete additivity.[2] When Σ is the power set $P(\Omega)$ of Ω (no non-measurable sets) countable additivity implies complete additivity unless the cardinality of Ω is as big as the least measurable cardinal, a situation rarely encountered in practice (see Appendix).

2.2 Bayes updating

Interpret pr as the credence function of a Bayesian agent. Suppose that the agent learns for certain that $F \in \Sigma$ is true where $pr(F) \neq 0$. How should her post-learning credence function pr' be related to her pre-learning credence function pr? A standard answer is that pr' is derived from pr by Bayes conditionalizing on F, i.e. $pr'(\bullet) = pr(\bullet / F) := \dfrac{pr(\bullet \cap F)}{pr(F)}$.

One possible justification for this updating rule (hereafter, Bayes updating) relies on

[2] Obviously Σ must have the appropriate closure properties for these requirements to make sense.

> *Prop 1.* Let (Ω, Σ, pr) be a classical probability space, and let $F \in \Sigma$ be such that $pr(F) \neq 0$. Then there is a unique functional $pr_F(\bullet)$ on Σ such that (a) $pr_F(\bullet)$ is a probability measure on Σ, and (b) for all $E \in \Sigma$ such that $E \subseteq F$, $pr_F(E) = pr(E)/pr(F)$.[3]

If properties (a) and (b) of Prop. 1 capture desirable features of updating when F is learned with certainty then Bayes updating is uniquely singled out since $pr_F(\bullet)$ is just $pr(\bullet/F)$. It remains to give a convincing justification for affirming the antecedent.

The most widely discussed justification for Bayes updating is the Lewis-Teller diachronic Dutch book argument.[4] A bookie employs a two-stage suite of bets using the agent's personal probability pr as the fair betting quotient.

> *Stage 1.* The bookie sells the agent an unconditional bet on F with betting quotient $pr(F)$ and stakes S_F.[5] At the same time the bookie buys from the agent a bet on E conditional on F using the betting quotient $pr(E/F) = pr(E \cap F)/pr(F)$ and stakes $S_{E/F}$. ('Conditional on F' means that if F is found to be false, the bet is called off.) The truth value of F is then ascertained and the result is announced to both the agent and the bookie. If F is found to be false the bookie collects on the unconditional bet, calls off the conditional bet and closes shop. If F is found to be true the bookie proceeds to the next stage.
>
> *Stage 2.* The bookie sells the agent an unconditional bet on E with stakes S_E and betting quotient given by the agent's probability pr' of E updated on the knowledge that F is true. The truth of E is ascertained and the remaining bets are settled.

If the agent's updated pr' is not equal to the conditional probability $pr(E/F)$ then the stakes of the three bets can be chosen so that, come what may, the agent loses money; conversely, if $pr'(E) = pr(E/F)$ then the agent is protected from diachronic Dutch book.

---

[3]For a proof see Cassenelli and Zanghi (1983).

[4]See Teller (1976). For an overview of Dutch book argumentation see Vineberg (2016).

[5]This bet is a contract whereby the agent agrees to pay the bookie $pr(F) \cdot S_F$ in order to collect S_F from the bookie if F is true and nothing is F is false.

This style of justification has been subjected to a number of criticisms. In particular, it has been claimed that diachronic Dutch book provides merely pragmatic grounds for adopting Bayes updating and does not show that an agent who departs from Bayes updating will necessarily have irrational credences (see Christensen 1991). More disturbingly, Gallow (2019) shows that Bayes updating is itself diachronically Dutch bookable in a scenario where an agent can acquire certainty about one of a pair of events $F_1, F_2 \in \Sigma$ that do not form a partition but overlap in a certain way.[6]

While I share qualms about the attempted justifications of Bayes updating I set them aside in the present context. Here my concern is not with justifying Bayes updating but rather with its impotence when faced with zero probability events.

3 Simple Bayesianism for quantum probability

I begin by emphasizing that what is at issue here is *not* what its proponents dub quantum Bayesianism (or QBism), the view that *all* quantum probabilities are to be given a personalist interpretation.[7] Even if, contrary to the QBians, the probability calculated from QM for, say, the decay of a hydrogen atom represents an objective chance, there is still the matter of how quantum physicists make inferences about the behavior of quantum systems. I assume for present purposes that these inferences are to be treated in the framework of Bayesian personalism, suitably adjusted to take account of the differences between classical and quantum event spaces.

3.1 Quantum probability

Quantum probability is the study of quantum probability measures on the projection lattice $\mathcal{P}(\mathfrak{N})$ of a von Neumann algebra $\mathfrak{N}$ acting on a Hilbert space $\mathcal{H}$ (see Hamhalter 2003). A projection E is an "observable," i.e. a self-adjoint operator, and it is idempotent, i.e. $E^2 = E$. $\mathcal{P}(\mathfrak{N})$ is equipped with a natural partial order, viz. for $E, F \in \mathcal{P}(\mathfrak{N})$, $F \leq F$ iff $range(E) \subseteq range(F)$. $\mathcal{P}(\mathfrak{N})$

[6] F_1 and F_2 overlap if $F_1 \cap F_2 \neq \emptyset$.

[7] For a readable introduction to QBism see von Baeyer (2016). For a critical assessment see Earman (2019a).

is closed under meet $\wedge$ (the least upper bound) and join $\vee$ (greatest lower bound). Complementation in $\mathcal{P}(\mathfrak{N})$ is understood as orthocomplementation, i.e. $E^c := E^\perp = I - E$. The elements of $\mathcal{P}(\mathfrak{N})$ are variously referred to as events, propositions, or Yes-No questions. It is assumed that it is in principle possible to settle these questions by appropriate measurements.

A quantum probability measure on $\mathcal{P}(\mathfrak{N})$ is a map $\Pr : \mathcal{P}(\mathfrak{N}) \to [0, 1]$ satisfying the quantum analogs of the classical probability axioms

(Qi) $\Pr(I) = 1$ (I the identity projection)

(Qii) $\Pr(E_1 \vee E_2) = \Pr(E_1) + \Pr(E_2)$ whenever $E_1, E_2 \in \mathcal{P}(\mathfrak{N})$ are mutually orthogonal.[8]

As with classical probability the additivity property (Qii) can be strengthened to countable and complete additivity.

Here I concentrate on the case of ordinary nonrelativistic QM where $\mathfrak{N} = \mathfrak{B}(\mathcal{H})$, the von Neumann algebra of all bounded operators acting on $\mathcal{H}$. When $\mathcal{H}$ is separable any family of mutually orthogonal projections is countable and, thus, complete additivity reduces to countable additivity, which in turn reduces to finite additivity when $\dim(\mathcal{H})$ is finite. A separable $\mathcal{H}$ suffices for most applications of QM, but it is easy to imagine cases where a non-separable $\mathcal{H}$ is required (see Section 4 below). However, unless $\dim(\mathcal{H})$ is as large as the least measurable cardinal countable additivity on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ implies complete additivity (see Appendix). To keep matters as simple as possible I focus for the most part on the case of a separable $\mathcal{H}$.

3.2 From quantum states to quantum probabilities and back

A quantum state ω is a complex valued, normed, positive linear functional on $\mathfrak{B}(\mathcal{H})$. Any such ω induces a quantum probability measure on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$, viz. $\Pr^\omega(E) := \omega(E)$, $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$, satisfies the properties (Qi) and (Qii) of the preceding section. If one takes the point of view that quantum states codify objective, observer independent features of quantum systems then, contra the QBians, the probabilities they induce are objective. But to repeat, this is a dispute that need not detain us.

[8] E_1 and E_2 are mutually orthogonal iff $E_1E_2 = E_2E_1 = O$ (the null projection). When E_1 and E_2 are mutually orthogonal $E_1 \vee E_2 = E_1 + E_2$.

A special class of states will occupy our attention here. Normal states are those that admit a density operator representation or, equivalently, are completely additive over families of mutually orthogonal projections.[9] Such states are the only states used in standard texts on QM, the implicit assumption being that only these states are physically realizable. A strong but not conclusive case can be made for this assumption (see Ruetsche 2011). It will be taken on board in the present discussion. A normal (respectively, non-normal) state induces a completely additive (respectively, merely finitely or merely countably additive) probability on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$.

Gleason's theorem provides an almost general converse:

> *Gleason*: Let Pr be a quantum probability on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ where $\dim(\mathcal{H}) > 2$ and $\mathcal{H}$ is separable. Then Pr has a unique extension to a state on $\mathfrak{B}(\mathcal{H})$ which is normal (respectively, non-normal) if Pr is countably additive (respectively, merely finitely additive).[10]

Later work showed that when $\mathcal{H}$ is non-separable a completely additive probability on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ has a unique extension to a normal state.

Gleason's theorem provides the crucial link in defining quantum updating.[11] But before turning to this matter a word needs to be said about how to read quantum probability statements.

3.3 Reading quantum probabilities

The formal similarity between the axioms of classical and quantum probability gives no hint at the deep differences that arise from the differences in the classical and quantum event spaces. To cite one example, when $\dim(\mathcal{H}) > 2$ there are no dispersion free probability measures on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$. By contrast

[9] See Kadison and Ringrose (1991), Vol. 2, Theorem 7.1.12. The density operator representation allows expectation values to be calculated via the trace prescription, viz. $\omega(A) = Tr(\varrho_\omega A)$, $A \in \mathfrak{B}(\mathcal{H})$, where ϱ_ω is the density operator corresponding to the normal state ω.

[10] When $\dim(\mathcal{H}) = 2$ there are probability measures on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ that do not extend to a state on $\mathfrak{B}(\mathcal{H})$. A detailed treatment of Gleason's theorem and its generalizations can be found in Hamhalter (2003).

[11] Gleason's theorem also provides the QBians with the means to claim that quantum states are merely representational devices used to keep track of quantum probability measures which, of course, they give a personalist reading.

if $\mathfrak{N}$ is an abelian von Neumann algebra there are dispersion free probability measures galore on $\mathcal{P}(\mathfrak{N})$; indeed, every pure state on an abelian $\mathfrak{N}$ induces a dispersion free measure. The absence of dispersion free measures on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ implies that there are no truth value assignments to $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ satisfying minimal natural requirements.[12] Since the elements of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ cannot be assumed to have simultaneous truth values, '$\Pr(E) = p$', for $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$, cannot be read as 'The probability that E is true equals p'. The alternative reading adopted here is 'The probability that a Yes-No measurement of E will return a Yes answer equals p', where there is no presumption that prior to the measurement E has a definite but unknown truth value.

3.4 Lüders updating

Suppose that a quantum Bayesian agent with an initial credence function Pr does a Yes-No experiment on $F \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$, where $\Pr(F) \neq 0$, and obtains a Yes answer. How should she update her credences in light of this new knowledge? One natural answer is that she should update by the closest quantum analog of Bayes conditionalization. Using '//' to stand for the sought-after quantum conditionalization, the requirement would be

> (†) If Pr is a quantum probability on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ and $F \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ is such that $\Pr(F) \neq 0$ then quantum conditionalization $\Pr(\bullet//F)$ on F is a quantum probability on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$, and $\Pr(E//F)$ agrees with classical conditionalization $\Pr(E/F) := \dfrac{\Pr(EF)}{\Pr(F)}$ for all $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ such that $EF = FE$, i.e. $\Pr(E//F) = \Pr(E/F)$.

[12] A truth value assignment is a map $V : \mathcal{P}(\mathfrak{B}(\mathcal{H})) \to \{True, False\}$. The following constraints are sufficient to generate the no-go result.

> (α) $V(I) = True$
>
> (β) For any mutually orthogonal $E_1, E_2 \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$, if $V(E_1) = True$ then $V(E_2) = False$
>
> (γ) For any mutually orthogonal $E_1, E_2 \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$, $V(E_1 \vee E_2) = True$ if either $V(E_1) = True$ or $V(E_2) = True$, and $V(E_1 \vee E_2) = False$ if both $V(E_1) = False$ and $V(E_2) = False$.

If such an assignment existed then $\Pr : \mathcal{P}(\mathfrak{B}(\mathcal{H})) \to \{1, 0\}$, where $\Pr(E) = 1$ if $V(E) = True$ and $\Pr(E) = 0$ if $V(E) = False$, would define a dispersion free probability.

If $E \leq F$ then $EF = FE = E$. So by (†) when $E \leq F$ and $\Pr(F) \neq 0$

$$\Pr(E//F) = \Pr(E/F) := \frac{\Pr(EF)}{\Pr(F)} = \frac{\Pr(E)}{\Pr(F)}.$$

Then appeal can be made to the quantum analog of Prop. 1:

> *Prop. 2*. Let Pr be a countably additive quantum probability measure on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ for separable $\mathcal{H}$ with $\dim(\mathcal{H}) > 2$, and let $F \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ be such that $\Pr(F) \neq 0$. Then there is a unique functional $\Pr(\bullet//F)$ on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ such that (a) $\Pr(\bullet//F)$ is a quantum probability, and (b) for all $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ such that $E \leq F$, $\Pr(E//F) = \dfrac{\Pr(E)}{\Pr(F)}$ (Bub (1977))

Castinelli and Zanghi (1983) show that Prop. 2 can be generalized to cover arbitrary von Neumann algebras. In addition, there is no problem in extending Prop. 2 to cover a non-separable $\mathcal{H}$ when the countable additivity of Pr is strengthened to complete additivity.

What is this unique functional $\Pr(\bullet//F)$, known in the literature as Lüders conditionalization? When Gleason's theorem applies and there is a unique extension of Pr to a normal state ω on $\mathfrak{B}(\mathcal{H})$, $\Pr(\bullet//F) = \dfrac{\omega(F \bullet F)}{\omega(F)} = \dfrac{\omega(F \bullet F)}{\Pr(F)}$. When $E, F \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ commute, $FEF = EF = FE = E \wedge F \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ and, thus, Lüders conditionalization can be expressed as $\Pr(E//F) = \dfrac{\Pr(EF)}{\Pr(F)}$, which agrees with classical conditionalization. Note that when E and F do not commute the numerator $\omega(FEF)$ in the expression for $\Pr(E//F)$ cannot be written as $\Pr(FEF)$ since $FEF \notin \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ and, hence, its probability is not defined.

When $\dim(\mathcal{H}) = 2$ all states are normal, but Gleason's theorem does not apply since there are quantum probability measures on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ that do not extend to any state on $\mathfrak{B}(\mathcal{H})$. Whether or not there is a plausible conditionalization rule for such measures is an interesting issue that will not be tackled here.

Lüders updating can also be motivated by adapting the Lewis-Teller diachronic Dutch book construction to quantum probabilities if bets on elements of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ are settled by making the appropriate Yes-No measurements and if $\Pr(E//F)$ is used to set the agent's Stage 1 fair betting quotient for a bet on E conditional on F, the understanding being that the bet is called off if a Yes-No measurement of F returns a No answer. And the construction shows that the stakes for the Stage 1 and 2 bets can be chosen so that the agent is Dutch booked if and only if her probability of E, updated on the knowledge that a Yes-No measurement of F yielded a Yes answer, fails to equal $\Pr(E//F)$ (see Earman (2019b)).

Needless to say, the qualms about the diachronic Dutch book argument for classical probability carry over to quantum probability. But for present purposes I am going to conduct the discussion below on the presumption that updating on non-zero probability events for classical and quantum probabilities follow respectively Bayes and Lüders updating. The issue is then is how to proceed when zero-probability events are encountered.

4 Regularity and strict coherence

The problem for both Bayes updating for classical probability and Lüders updating for quantum probability arises from the fact that classical conditionalization $pr(\bullet/F)$ and Lüders conditionalization $\Pr(\bullet//F)$ are both undefined when the the initial probabilities $pr(F)$ and $\Pr(F)$ of the conditioning event are zero. The problem cannot be ignored since agents can and do learn facts to which they initially assign zero credence.

The plight of the would-be Bayesian agent. 'I want to be a good Bayesian. Towards this end I have a classical (respectively, quantum) credence function that satisfies the axioms of classical probability (respectively, quantum probability). It assigns zero credence to the event F. But I just verified by direct observation (or by experiment, or by reliable testimony) that F has occurred. How should I revise my credences if I am to continue as a good Bayesian? I cannot use Bayes updating (respectively, Lüders updating) since it is undefined. Please help—I want to be good, but I don't know what a good Bayesian is supposed to do in the face of this challenge.'

I will discuss various forms of tough love advice suggested in the literature, beginning with the requirement of regularity/strict coherence

Tough love (version 1). 'I commend you for the fact that your initial credences conform to the axioms of probability. Your credences are coherent, meaning that they are immune from being Dutch booked when they are used to set fair betting quotients. However, I have to tell you that you got off on the wrong foot—you did not start with a credence function that fully conforms to the requirements of rationality of belief. That is because your initial credences, though coherent, are *not strictly coherent*: if they are used to set fair betting quotients then there is a family of bets each of which you find fair or favorable but collectively have the feature that in no possible case do you have a positive gain and in some possible case you suffer a loss. To be strictly coherent your credences not only have to conform to the axioms of probability as discussed above but they must be regular; that is, in the case of classical probability a credence of zero is assigned only to the null event and in the case of quantum probability only to the null projection. A number authorities have promoted regularity as a rationality requirement (see Kemeny 1955, Shimony 1955, Lewis 1980, Skyrms 1980, Jackson 1987, Jeffrey 1992). So my advice to you is to start over and adopt a regular/strictly coherent credence function. Good luck!'

Tough love is sometimes required. But in this case there are various reasons to be leery of embracing it. In this section I will mention the most pressing one. Others will emerge in the discussion below. Start with the truism that ought implies can. But the relevant 'can' here may be missing. In particular, the injunction of regularity/strict coherence cannot be satisfied if there are more than a countable number of mutually exclusive alternatives to which probability is assigned. (The proof is well known, but to save readers from having to look it up I sketch it in the Appendix.) In the case of a classical probability space this means that the set Σ of measurable sets contains more than a countable number of mutually disjoint sets; in the case of a quantum probability it means that the projection lattice $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ contains more than a countable number of mutually orthogonal projections, which is the case iff $\mathcal{H}$ is non-separable.

An example from classical probability where the 'can' is missing is given by an infinitely sharp dart thrown at dart board isomorphic to a unit interval of the real line; or if the idealization of an infinitely sharp dart is rejected, count as a "hit" the coincidence of some chosen geometrical point on the tip of the dart with a geometrical point on the dart board. There are a power of continuum $\mathfrak{c}$ mutually exclusive outcomes of a toss of the dart, so if a probability is to be assigned to each outcome, Σ must have at least $\mathfrak{c}$ mutually disjoint sets. An example from quantum probability is given by an infinite spin chain consisting of a countably infinite number of spin sites, each of which may be in a state of spin up or spin down. A Hilbert space of dimension $2^{\aleph_0}$, which is equal to $\mathfrak{c}$ if the continuum hypothesis is correct, is needed. Consequently, there are $\mathfrak{c}$ mutually orthogonal projections. In both cases a real-valued probability measure can give non-zero probability to only a countable subset of the alternatives.

Rather than give up on regularity its advocates sometimes propose to change one of the basic rules of the game.

5 Regularity redux: infinitesimal probabilities

Tough love (version 2). 'The previous advice I gave you was both too tough and not tough enough. It was too tough because I asked you to do what in some cases is impossible. But it was not tough enough because I left the impression that it was ok for you proceed using standard probability theory in which probabilities are real-valued. In order to achieve regularity without running afoul of ought-implies-can you should, if necessary, adopt a version of probability theory that fattens the range of the probability function (to use Brian Skyrms' expression).

'The most discussed version of this strategy is to allow probabilities to take hyperreal values; with this fattening, more than a countable number of mutually exclusive events can be assigned non-zero probabilities because they can be assigned infinitesimal values, such a value being greater than zero but less than $1/N$ for any positive integer N. The work of logicians and mathematicians have made talk of the infinitesimal respectable by providing rigorous accounts of non-Archimedean extensions of the real numbers $\mathbb{R}$. For example, Abraham Robinson showed that there are non-Archimedean elementary extensions of the reals that preserve all the properties of the reals that can be stated in a first-order language. Using this work Bernstein and Wattenberg (1969) showed how to treat the dart board example so as to avoid the ought-implies-can objection. In their construction the probability assigned to (the singleton set of) any real number in the unit interval of $\mathbb{R}$ is a fixed infinitesimal, and any non-empty set of Lebesgue measure zero gets positive infinitesimal measure. Further, their measure of any Lebesgue measurable set differs from standard Lebesgue measure by an infinitesimal. This is a model for you to emulate.'

Here the tough love sermonette must be interrupted to deal with the worry that while hyperreals may overcome specific cases of the ought-implies-can objection, they will not provide a general resolution. Pruss (2013) showed that no matter what non-Archimedean extension $\mathbb{R}^*$ of the real numbers $\mathbb{R}$ is chosen there is an event space Ω and an algebra of subsets Σ such that there is no regular probability measure from Σ into $\mathbb{R}^*$—intuitively, fattening the domain of the probability measure can outrun the ability of the chosen $\mathbb{R}^*$-fattening of the range to deliver regularity.[13] Hofweber (2014) counters, in effect, that one size need not fit all; that is, the choice of the non-Archimedean extension $\mathbb{R}^*$ to fatten the range of the probability function need not be made once and for all but may be tailored to fit the domain. Given this flexibility, achieving regularity is always achievable by a hyperreal fattening of the range:

> *Prop. 3* (Hofweber 2014, Hofweber and Schindler 2016). For any sample space Ω and any algebra Σ of subsets of Ω there is a hyperreal field $\mathbb{R}^*$ of at most size $2^{|\Omega|}$ and a regular probability measure from Σ into $\mathbb{R}^*$.

With this flexible approach to fattening the range of the probability function the way is open to formulate a conjecture that would make tough love version 2 more palatable and more helpful. Towards this end note that if the hyperreal number $r^* \in \mathbb{R}^*$ is finite (i.e. there is a standard real $r > 0$ such that $-r < r^* < r$) then there is a unique standard real $st(r^*)$ closest to r^* (i.e. $st(r^*) = \inf\{r \in R : r^* \leq r\}$) such that $|st(r^*) - r^*|$ is an infinitesimal. This $st(r^*)$ is referred to as the standard part of r^*. Let (Ω, Σ, pr) be a standard classical probability space. Call the probability space (Ω, Σ', pr^*) where pr^* takes values in the hyperreal field $\mathbb{R}^*$ a regular hyperreal extension of (Ω, Σ, pr) iff $\Sigma' \supseteq \Sigma$ and for all $E \in \Sigma$, $pr^*(E) \neq 0$ and $st(pr^*(E)) = pr(E)$. (The Bernstein and Wattenberg 1969 extension of Lebesgue measure is an example.)

> *Conjecture 1.* For any standard probability space (Ω, Σ, pr) there is a regular hyperreal extension (Ω, Σ', pr^*).

A version of Conjecture 1 for finitely additive probabilities is demonstrated by Hofweber and Schindler (2016), but it remains to be proven, or refuted, for countably or completely additive pr.

[13] Pruss' argument uses the axiom of choice and finite additivity.

[Aside: How to understand the concepts of countable and complete additivity for probability measures taking values in the hyperreals remains to be specified. The standard reals have the least upper bound property: a bounded non-decreasing sequence of reals has a least upper bound. So if pr is a real valued probability and $\{E_a\} \in \Sigma$ is a family of mutually disjoint events an infinite sum $\sum_a pr(E_a)$ can be understood as the least upper bound on finite partial sums. But the least upper bound property fails for the hyperreals. There are various proposals for how to understand convergence for a sequence of hyperreals, but as far as I am aware there is no general agreement on which proposal is correct or even on what correctness here means apart from consistency and fruitfulness. However, how to understand countable and complete additivity for hyperreal valued probabilities is an issue that need not detain us. The point is that Conjecture 1 remains to be proved for countably and completely additive real valued pr, and the additivity properties of a hyperreal valued pr^* need not concern us here. If (Ω, Σ', pr^*) is a hyperreal extension of (Ω, Σ, pr) then, by definition, the standard part $st(pr^*(\bullet))$ of $pr^*(\bullet)$ satisfies the same form of additivity as $pr(\bullet)$ *on the latter's domain* Σ though not necessarily on the former's domain $\Sigma' \supseteq \Sigma$. This is illustrated by the Bernstein and Wattenberg (1969) non-standard measure μ^* that extends Lebesgue measure μ on $[0,1] \subset \mathbb{R}$ to all subsets of $[0,1]$. The standard part $st(\mu^*(\bullet))$ of μ^* agrees with standard Lebesgue measure on all Lebesgue measurable sets and, thus, is countably additive over these sets. But, if the continuum hypothesis is correct, $st(\mu^*(\bullet))$ cannot be countably additive on the power set of the unit interval since then it would also be completely additive (see Appendix), producing a contradiction since $st(\mu^*(\{r\})) = 0$ for all $r \in [0,1]$, but if $st(\mu^*(\bullet))$ is completely additive on all subsets of $[0,1]$ then $st(\mu^*(\cup_{r\in[0,1]}\{r\})) = st(\mu^*([0,1])) = 1$.]

After this interruption the tough love sermonette can now continue.

Tough love (version 2.1). 'Abandoning your standard probability measure pr in favor of a regular hyperreal valued measure pr^* need not be as wrenching as you might have imagined. If Conjecture 1 is true you can *almost* retain all of your pr degrees of belief by choosing a regular hyperreal extension pr^*; for $|pr^*(E) - pr(E)|$ is an infinitesimal for all $E \in \Sigma$ and, thus, only infinitesimal adjustments in your real valued credences are needed to achieve regularity. Of course, if there is a hyperreal valued extension of pr there will be many. I cannot tell you which one to choose.'

For tough love version 2.1 to be viable it must be extended to quantum probabilities. This would require developing a non-standard measure theory for quantum probabilities wherein quantum probabilities are allowed to take on hyperreal values; developing a notion of quantum states as linear functionals on $\mathfrak{B}(\mathcal{H})$ that take values in the hypercomplex numbers and that induce hyperreal valued probabilities on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$; and proving a Gleason type theorem, showing that a hyperreal valued quantum measure satisfying appropriate additivity conditions (whatever this would mean) can be uniquely extended to a normal hypercomplex valued quantum state (whatever normality here would mean). Additionally, a quantum analog of Conjecture 1 would need to be formulated and proved. This is terra incognita and potentially a fruitful area for research, but it is not a project that can be pursued here. So I return to non-standard measure theory for classical probability. But I emphasize that until this research program has borne fruit the idea that the problem at issue can be avoided by the device of hyperreal probabilities is in danger of foundering in the quantum domain.

6 Why regularity (strict coherence) is not a good regulatory ideal

6.1 Weak motivation

Although regularity/strict coherence—achieved, if necessary, with the help of infinitesimal probabilities—has been widely promoted as a rationality constraint on credences, the rationales offered are weak in comparison with the rationales for simple coherence. Consider, for example Dutch book argument for simple coherence, which can be given either a pragmatic or a non-pragmatic spin. On the pragmatic reading the Dutch book argument provides a strong nudge towards conforming your credences to the dictates of probability by showing that a failure to conform means that a clever bookie can clean out your bank account if you use your credence function as a fair betting quotient, while the converse Dutch book argument shows that conformity confers immunity to such sure ruin. By contrast, if coherence holds the failure of regularity (or strict coherence) means only that you are indifferent to or find favorable a family of bets, the net result of which is that in no case do you have a gain while in some case you have a loss. The surety of a loss in some case is due to the fact you are willing to buy a bet on a measurable event $E \neq \emptyset$ where you have no gain if E is true but you lose the stakes of the bet if E is false. But this should not be be too worrisome for you since you think that the probability of losing the stakes is zero.

On the non-pragmatic reading of Dutch book what the argument for simple coherence (supposedly) reveals is that a that failure to conform your credences to the dictates of probability results in a structural incoherence in your degrees of belief and preferences: you are indifferent to or find favorable each of a family of bets but at the same time prefer not betting on the family. By contrast, the failure to obey regularity/strict coherence reveals no comparable structural incoherence in your degrees of belief and preferences. Note also that non-Dutch book rationales for simple coherence, such as scoring rule arguments, do not serve as rationales for strict coherence.

6.2 Cases where regularity for credence functions is neither desirable nor maintainable

As already noted, for most applications of ordinary QM a separable Hilbert space $\mathcal{H}$ suffices, and for such cases any family of mutually orthogonal elements of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ is countable. Hence, real valued probability measures on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ can be regular, and there is no need to resort to infinitessimal probabilities or the like. However, *every* normal pure state ψ induces on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ a real valued probability measure $\Pr^{\psi}$ that is non-regular. So if, contrary to the QBians, normal pure states induce objective chances, then these chances are non-regular. And if rational credence should strive to match objective chance, then regularity is not a desirable regulatory ideal for credence.

In some instances Lüders updating automatically makes rational credence match objective chance. Specifically, suppose that $\mathcal{H}$ is separable and $\dim(\mathcal{H}) > 2$, and let Pr be a countably additive and regular credence function on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$. Lüders updating Pr on what is called the support projection S_{ψ} of a normal pure state ψ gives the following result, purely as a theorem of quantum probability: If $\Pr(S_{\psi}) \neq 0$ then $\Pr(\bullet//S_{\psi}) = \psi(\bullet)$ (see Section 10 below). The updated credence function matches the non-regular chance function $\psi(\bullet)$, and in this instance the non-regularity is arguably a desirable feature rather than something to avoided or smoothed over by filling in zero probabilities with infinitesimal. Furthermore, the fleeting nature of regularity in this instance is quite a general phenomenon.

6.3 The ephemeral nature of regularity

Regularity was put forward in order to avoid the problem at issue. But the avoidance doesn't last long. I will illustrate for classical real valued probabilities. Suppose that our would-be good Bayesian agent adopts an initial credence function pr that is regular. Consider what happens when this agent learns with certainty that $F \in \Sigma$ is true. Except in the trivial case where $\Sigma = \{\Omega, \emptyset\}$ we may assume that $F \neq \Omega$ and that our agent's initial credence function pr is such that $pr(F) < 1$ (otherwise he hasn't learned anything of which he wasn't probabilistically certain to begin with). Since our agent's initial credence function is regular, $pr(F) > 0$ and, thus, upon learning that F is true he updates by Bayes conditionalization to $pr'(\bullet) := \dfrac{pr(\bullet \cap F)}{pr(F)}$. Our agent's updated credence function is non-regular—it assigns zero probability to any $E \in \Sigma$ disjoint from F. Regularity is an utterly ephemeral property for real valued probability, and it is no different when the range of the probability function is fattened to include hyperreals. Nor is regularity any less ephemeral for quantum probabilities under Lüders updating.

The ephemeral nature of regularity calls into question its status. If an agent's initial credence function counts as rational and she Bayes (or Lüders) updates then her updated credence function should count as rational; but, if so, regularity cannot be a rationality requirement *tout court* since it can be lost in updating. Perhaps regularity can be maintained as a rationality constraint on initial credences, and thereafter retained as necessary companion of Bayes (or Lüders) updating. Leaving aside the question-begging nature of the retention, there is the problem that the retention requires re-regularization of the agent's credence function after each updating, and our would-be good Bayesian will want to know *which* re-regularized credence function she ought to choose and why. She is still waiting for an answer.

A way to escape this difficulty is to ensure that an agent's learning experiences are always uncertain, shifting her degrees of belief in non-null elements of Σ (or $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$) from one non-extreme value to another but never to 0 or 1. How to model such learning will be discussed below in Sections 8 and 9, but before going there let us consider another tactic for coping with zero probability events.

7 Popper functions/full conditional probability

Tough love (version 3). 'The source of your problems is your simple-minded form of Bayesianism which takes unconditional probability as the basic probability concept. You should follow the lead of Karl Popper and take conditional probability as the fundamental concept. Philosophers usually discuss this development in terms of what have come to be called Popper functions (see Popper (1961), Appendices iv and v). I will recommend a special case of Popper functions that mathematicians call full conditional probabilities. The conditional probability $pr(\bullet \wr \circ)$ of a classical full conditional probability space $(\Omega, \Sigma \text{ x } \Sigma^0, pr(\bullet \wr \circ))$ is a map from $\Sigma \text{ x } \Sigma^0$ to $[0, 1]$ where Σ^0 consists of the non-null elements of Σ. It is required to satisfy

(a) $pr(\bullet \wr G)$ is a probability measure on Σ for all $G \in \Sigma^0$
(b) $pr(G \wr G) = 1$ for all $G \in \Sigma^0$
(c) $pr(E \wr G) = pr(E \wr F)pr(F \wr G)$ for $E \subseteq F \subseteq G \subseteq \Omega$ and $F \neq \emptyset$.[14]

(A Popper function has domain $\Sigma \text{ x } \widetilde{\Sigma}^0$ where $\widetilde{\Sigma}^0$ is a subset of Σ^0 that must satisfy some finicky closure properties. When $\widetilde{\Sigma}^0$ is a proper subset of Σ^0 Popper functions are not well suited to help with the updating problem since that application requires that any member of Σ^0 can be plugged into the right hand slot of $pr(\bullet \wr \circ)$.) My advice to you is to model learning from experience using full conditional probabilities. Details to follow.'

Here the tough love advice is interrupted to provide more background. If $pr(\bullet)$ is an unconditional probability measure of a standard (real-valued) probability space $(\Omega, \Sigma, pr(\bullet))$ say that the full conditional probability $pr(\bullet \wr \circ)$ in $(\Omega, \Sigma, pr(\bullet \wr \circ))$ extends $pr(\bullet)$ just in case $pr(E) = pr(E \wr \Omega)$ for all $E \in \Sigma$. Note that if $pr(\bullet \wr \circ)$ extends $pr(\bullet)$ and $pr(F) \neq 0$ then $pr(E/F) := \dfrac{pr(E \cap F)}{pr(F)} = pr(E \wr F)$ for all $E \in \Sigma$. One can conjecture:

> *Conjecture 2.* For any standard classical probability space (Ω, Σ, pr) there is a full conditional extension $(\Omega, \Sigma, pr(\bullet \wr \circ))$.

[14] In the presence of (a) and (b) condition (c) is equivalent to

$$pr(E \cap F \wr G) = pr(F \wr G)pr(E \wr F \cap G) \text{ for } E, F \in \Sigma,\ G \neq \emptyset, \text{ and } F \cap G \neq \emptyset.$$

Versions of Conjecture 2 were proven by Krauss (1968) and Dubins (1975) for finitely additive probabilities, but as far as I am aware Conjecture 2 remains to be proved for countably and completely additive probabilities. Since extensions, when they exist, are not unique the correspondence between an unconditional probability and full conditional extensions is one-many.

There is also a close relation between full conditional probabilities and regular hyperreal probabilities. If pr^* is a regular hyperreal valued probability on Σ then $pr(E \wr F) := st(pr^*(E/F))$, $E \in \Sigma$ and $F \in \Sigma^0$, defines a full-conditional probability on Σ x Σ^0. Thus, a proof of Conjecture 1 would also give a proof of Conjecture 2. There are also results showing how to go in the other direction, from full conditional probabilities to regular hyperreal valued probability measures (see McGee 1994 and Halpern 2009). This two-way traffic has led to the claim that the two approaches amount to the same thing, but Halpern (2009) cautions that such a claim needs to be carefully qualified.[15]

A parallel development for quantum probability would begin by defining a full conditional quantum probability $\Pr(\bullet \wr \circ)$ as a map from $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ x $\mathcal{P}(\mathfrak{B}(\mathcal{H}))^0$ to $[0,1]$, where $\mathcal{P}(\mathfrak{B}(\mathcal{H}))^0$ consists of all non-null projections. $\Pr(\bullet \wr \circ)$ is required to satisfy the analogs of (a)-(c):

(a′) $\Pr(\bullet \wr G)$ is a probability measure on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ for all $G \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))^0$

(b′) $\Pr(G \wr G) = 1$ for all $G \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))^0$

(c′) $\Pr(E \wr G) = \Pr(E \wr F)\Pr(F \wr G)$ for $E \leq F \leq G \leq I$ and $F \neq O$.

In analogy with classical probability say that the full conditional quantum probability $\Pr(\bullet \wr \circ)$ extends the unconditional $\Pr(\bullet)$ just in case $\Pr(\bullet \wr I) = \Pr(\bullet)$. It would need to be proved that if $\Pr(\bullet \wr \circ)$ extends $\Pr(\bullet)$ and $\Pr(F) \neq 0$, $F \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$, then $\Pr(\bullet \wr F) = \Pr(\bullet // F)$. Also required is a notion of conditional quantum state on $\mathfrak{B}(\mathcal{H}))$ x $\mathfrak{B}(\mathcal{H})^0$ that induces on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ x $\mathcal{P}(\mathfrak{B}(\mathcal{H}))^0$ a full conditional quantum probability. And in addition a Gleason type theorem relating full conditional quantum probabilities to conditional

[15] Actually, the alleged equivalence of the two approaches is usually discussed in the literature in terms of Popper functions rather than full conditional probabilities, which I think is a mistake.

quantum states would need to be proved. How to provide these ingredients is far from obvious, so let us concentrate on classical probabilities. But I emphasize that until these ingredients are supplied the idea that the problem of updating on zero probability events can be handled by the device of full conditional probabilities is in danger of foundering in the quantum domain.

Let us now return to the tough love lesson for classical probabilities.

Tough love (version 3, continued). 'Instead of starting with an initial unconditional probability $pr(\bullet)$ start with a full conditional $pr(\bullet \wr \circ)$; the associated unconditional probability can be defined as $pr(\bullet) := pr(\bullet \wr \Omega)$. Or if you insist on starting with an unconditional $pr(\bullet)$ extend it to a full conditional $pr(\bullet \wr \circ)$ (which you can always do if Conjecture 2 is correct). In either case if your unconditional $pr(\bullet)$ is irregular and $pr(F) = 0$ you cannot Bayes update on F by plugging F into the right hand slot of the Bayes conditioning $pr(\bullet/\circ)$ of $pr(\bullet)$. But not to worry. You can plug any $F \in \Sigma^0$ into the right hand slot of the full conditional $pr(\bullet \wr \circ)$ and then take $pr'(\bullet) := pr(\bullet \wr F)$ to be the F-updating of your unconditional probability.'

Now what? The suggested F-updating has produced a measure $pr'(\bullet)$ that is not regular, so the agent cannot Bayes update $pr'(\bullet)$ if he subsequently learns an $F' \in \Sigma$, $F' \neq \emptyset$, such that $pr'(F') = 0$. Nor can he find the help he needs in the conditional probability $pr'(\bullet \wr \circ) := pr(\bullet \wr \circ \cap F)$ since this is not a full conditional probability. Being full conditional is as ephemeral a property of conditional probabilities as regularity is for unconditional probabilities.

Of course, the agent can seek the help of another full conditional probability $pr''(\bullet \wr \circ)$, and if Conjecture 2 is correct he can limit his choice to a full conditional extension of $pr'(\bullet) = pr(\bullet \wr F)$. And the agent has to seek similar help anew with each new thing he learns. What is objectionable here is not the Sisyphean nature of the labors to which the tough love advice version 3 commits the would-be good Bayesian but rather the conceit that the resort to full conditional probabilities offers any real help on how to be a good Bayesian. The advice, 'Choose some full conditional probability', is empty advice since it is tantamount to the advice to choose one among possible belief revision strategies for dealing with zero probability events. Even if Conjecture 2 is true in great generality and the choice of full conditional probability is limited to an extension of the latest non-regular unconditional probability, the choice is far from unique. Which extension to choose and why is what the would-be good Bayesian wants to know. He is still waiting for an answer.

Time again to try a new tactic.

8 Uncertain learning

Suppose an agent's initial credences are codified in a classical probability space (Ω, Σ, pr). She has a learning experience and announces that as a result she has new degrees of belief $pr_{new}(F_a)$, $\sum_a pr_{new}(F_a) = 1$, for the elements of a partition $\{F_a\}$ of Σ. These new credences can be extended to a new probability measure on Σ by means of Jeffrey conditionalization if her conditional credences are rigid over the partition in the sense that $pr_{new}(E/F_a) = pr(E/F_a)$ for all elements F_a of the partition and all $E \in \Sigma$. The extension rule is given by:

$$pr_{new}(E) := \textstyle\sum_a pr(E/F_a) pr_{new}(F_a) \quad \text{for all } E \in \Sigma. \tag{JC}$$

If there is an $F \in \{F_a\}$ such that $pr_{new}(F) = 1$ then (JC) reduces to standard Bayes conditioning: $pr_{new}(E) = pr(E/F)$ for all $E \in \Sigma$.

The diachronic Dutch book argument for updating by Bayes conditionalization for certain learning has been extended to updating by (JC) for uncertain learning (see Armendt 1980 and Skyrms 1987). But the qualms about the diachronic Dutch book justification for Bayes updating apply here as well (see Gallow 2019).

A non-Dutch book justification can be obtained from the following generalization of Prop. 2:

> *Prop. 4.* Let (Ω, Σ, pr) be a classical probability space, and let $\{F_a\}$, $F_a \in \Sigma$, be a partition such that $pr(F_a) \neq 0$ for all F_a. Then there is a unique functional $pr_{\{F_a\}}(\bullet)$ on Σ such that (a) $pr_{\{F_a\}}(\bullet)$ is a probability measure on Σ, and (b) for all $E \in \Sigma$ such that $E \subseteq F_b$ for some $F_b \in \{F_a\}$, $pr_{\{F_a\}}(E) = \dfrac{pr(E)}{pr(F_b)} pr_{\{F_a\}}(F_b)$.

The unique functional is given by (JC). Thus, if clauses (a) and (b) of Prop. 4 capture a desirable feature of updating under uncertain learning over a partition then updating by Jeffrey conditionalization is uniquely singled out.

How, if at all, does Jeffrey conditionalization help with our main problem? The short answer is that by itself it doesn't help. As a result of a learning experience an agent may assign a positive credence $pr_{new}(F_{a^*}) > 0$ to some element F_{a^*} of the relevant partition, even though prior to the learning experience she assigned $pr(F_{a^*}) = 0$. But if $pr(F_{a^*}) = 0$ (JC) cannot be used to extend the newly acquired credences to a new probability measure on Σ since the term $pr(E/F_{a^*})$ in the sum on the rhs of (JC) is undefined. If this term is simply dropped from the sum then an inconsistency results since, by hypothesis, $pr_{new}(F_{a^*}) > 0$ whereas (JC) sans the term in question gives $pr_{new}(F_{a^*}) = 0$.

Although Jeffrey conditionalization cannot provide a positive solution to our problems it can contribute to negative one. If the agent's learning is truly uncertain in the sense that $0 < pr_{new}(F_a) < 1$ for all elements of the partition $\{F_a\}$ then (JC) does not make regularity a fleeting property as does ordinary Bayes conditioning; indeed, for uncertain learning (JC) preserves regularity in that if the initial $pr(\bullet)$ is regular then so is $pr_{new}(\bullet)$. Using this feature and some mixing and matching produces some new tough love advice.

Tough love (version 4). 'Using hyperreal valued probabilities if necessary, adjust your initial credences so that they are regular. Thereafter confine your learning experiences to partitions $\{F_a\}$ and to inquiry methods that ensure your experience will produce truly uncertain learning over the partition, i.e. $0 < pr_{new}(F_a) < 1$ for all elements of the partition, e.g. make optically unaided eyeball observations by dim, flickering candle light. And after each such learning experience update by (JC). If you follow this advice you will never have to struggle with how to update on zero probability events or wonder how an event with an initially zero credence can earn a positive credence.'

Following this tough love 'solves' our problem by not allowing it to arise in the first instance and then ensuring that it does not arise at a later stage by not allowing the agent's veil of ignorance to be entirely lifted on any proposition. Perhaps a problem avoided is a problem solved. But in this instance the instruments of avoidance are questionable. Enough has already been said about the dubious status of regularity as a rationality requirement. As for uncertain learning, allowing for it is a virtue but requiring it seems self-serving. The alternative is to claim that no artificial means are needed to produce truly uncertain learning and that, in fact, all learning from experience is truly uncertain—even under the most optimal conditions of observation one can never be entirely certain that any element of any relevant partition is true. Such a claim requires argument.

One might guess that the proper quantum analog of (JC) would be obtained by replacing Bayes conditionalized factors in (JC) by Lüders conditionalized factors, producing

$$\Pr_{new}(E) = \textstyle\sum_a \Pr(E//F_a) \Pr_{new}(F_a) \quad \text{for all } E \in \mathcal{P}(\mathfrak{B}(\mathcal{H})) \qquad \text{(QJC)}$$

where $\{F_a\} \subset \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ is now a family of mutually orthogonal projections such that $\sum_a F_a = I$, the quantum analog of a partition in classical probability. As with (JC) there is a proviso to the effect that $\Pr(F_a) > 0$ for all elements of the partition. However, (QJC) does not pass an obvious consistency condition; namely, if $\Pr_{new}(F_a) = \Pr(F_a)$ for all F_a of the partition then it should be the case that $\Pr_{new}(E) = \Pr(E)$. To see what the consistency condition requires assume that $\dim(\mathcal{H}) > 2$, that $\mathcal{H}$ is separable, and that Pr is countably additive (or alternatively that $\mathcal{H}$ is non-separable and Pr is completely additive). By Gleason's theorem Pr extends uniquely to a normal state ω. Since $\sum_a F_a = I$ we have

$$\Pr(E) = \omega((\textstyle\sum_a F_a)E(\sum_b F_b)) = \sum_a \omega(F_a E F_a) + \sum_{b \neq c} \omega(F_b E F_c)$$

Using the proviso that $\Pr(F_a) = \omega(F_a) > 0$ for all elements of the partition and the definition of '//',

$$\omega(F_a E F_a) = \frac{\omega(F_a E F_a)}{\omega(F_a)} \omega(F_a) = \Pr(E//F_a) \Pr(F_a).$$

Collecting these results gives

$$\Pr(E) \ = \textstyle\sum_a \Pr(E//F_a) \Pr(F_a) + \sum_{b \neq c} \omega(F_b E F_c).^{16}$$

The extra ω-term on the rhs represents interference effects, one of the characteristic features of QM resulting from the non-classical (= non-abelian) event structure.

To achieve consistency modify (QJC) to

$$\begin{aligned}\Pr_{new}(E) = \textstyle\sum_a \Pr(E//F_a) \Pr_{new}(F_a) & \\ + \sum_{b \neq c} \omega(F_b E F_c) \quad \text{for all } E \in \mathcal{P}(\mathfrak{B}(\mathcal{H})) &\end{aligned} \qquad \text{(QJC+)}$$

where the proviso that $\Pr(F_a) > 0$ for all elements of the partition is understood to be in effect. In addition to requiring rigidity of the Lüders conditional probabilities, (QJC+) also requires rigidity of the interference term,

[16] Note that in general $\omega(F_b E F_c)$ cannot be written as $\Pr(F_b E F_c)$ since if E and the F's do not commute $(F_b E F_c) \notin \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ and $F_b E F_c$ is not assigned a probability.

i.e. $\sum\limits_{b \neq c} \omega_{new}(F_b E F_c) = \sum\limits_{b \neq c} \omega(F_b E F_c)$ where ω_{new} is the normal state extending $\Pr_{new}$. The latter requirement limits the applicability of the quantum version of (JC) since, for example, Jeffrey updating from a Pr for which the interference term vanishes will not in general result in a $\Pr_{new}$ for which the ω_{new}-term vanishes.

It remains to be shown that the diachronic Dutch book argument for (JC) can be adapted to (QJC+) and/or that the quantum analog of Prop. 4 can be shown to hold for (QJC+). Even if the details do work out (QJC+) can be employed as part of a strategy for avoiding updating on zero-probability quantum events when $\mathcal{H}$ is non-separable only if the as yet non-existent apparatus for hyperreal valued quantum probabilities is developed.

9 Joining Popper with Jeffrey

Let $(\Omega, \Sigma, pr(\bullet))$ and $(\Omega, \Sigma \text{ x } \Sigma^0, pr(\bullet \wr \circ))$ be standard classical and full conditional classical probability spaces respectively with $pr(\bullet) = pr(\bullet \wr \Omega)$. An agent with initial unconditional probability $pr(\bullet)$ has a learning experience over partition $\{F_a\}$, $F_a \in \Sigma^0$, resulting in new probabilities $pr_{new}(F_a)$ for the elements of a partition. Joining Jeffrey and Popper in unholy matrimony, our agent updates $pr(\bullet)$ by Jeffrey-Popper conditionalization by using

$$pr_{new}(E) := \textstyle\sum_a pr(E \wr F_a) pr_{new}(F_a) \quad \text{for all } E \in \Sigma. \tag{JPC}$$

(JPC) satisfies the obvious consistency condition that when $pr_{new}(F_a) = pr(F_a)$ for all elements of the partition then $pr_{new}(E) = pr(E)$ for all $E \in \Sigma$.[17] In light of the discussion of the preceding section the quantum analog of (JPC) would have to make allowances for interference effects. But since the apparatus of full conditional probabilities for quantum probability theory remains to be developed I concentrate on classical probability.

If $pr(F_a) \neq 0$ for all elements of the partition then (JPC) reduces to (JC). And if, further, there is some element of the partition whose new probability is 1 then (JPC) reduces to Bayes conditionalization. The advantage of (JPC)

[17] Terms in the sum $\sum\limits_a pr(E \wr F_a) pr(F_a)$ for which $pr(F_a) = 0$ do not contribute and can be dropped. For the remaining terms $\sum\limits_a pr(E \wr F_a) pr(F_a) = \sum\limits_a pr(E/F_a) pr(F_a)$ $= \sum\limits_a pr(E \cap F_a) = pr(E)$, where the last inequality follows from the principle of total probability.

lies in the case where there is some element F_a of the partition for which $pr(F_a) = 0$. For in this case (JC) is stymied whereas (JPC) is well-defined and suffices to produce the updated $pr_{new}(\bullet)$. It seems then that, unlike (JC), (JPC) can both allow for uncertain learning and can contribute to a positive solution to our problem.

While this unholy Jeffrey-Popper marriage has some advantages it faces the by now familiar problems at the next iteration. Suppose that the agent has a second learning experience over the partition $\{G_a\}$ resulting in new-new credences $pr_{new-new}(G_a)$ for the elements of the partition. If Conjecture 2 is correct there is a full conditional $pr_{new}(\bullet \wr \circ)$ such that $pr_{new}(\bullet) = pr_{new}(\bullet \wr \Omega)$. She can then update to

$$pr_{new-new}(E) = \sum_a pr_{new}(E \wr G_a) pr_{new-new}(G_a) \quad \text{for all } E \in \Sigma.$$

The difficulty is that if there is one full conditional $pr_{new}(\bullet \wr \circ)$ such that $pr_{new}(\bullet) = pr_{new}(\bullet \wr \Omega)$ then there are many, except of course when $pr_{new}(\bullet)$ is regular and (JPC) reduces to (JC). Which of these $pr_{new}(\bullet \wr \circ)$ should be used to (JPC) update and why? The formalism recapitulates our problem but doesn't help to resolve it.

One could go on to detail troubles in arranging a quantum version of the Jeffrey-Popper marriage, but enough misery is enough. After all of the disappointments it is time for a success story. QM provides one, but it is a story in which regularity is trashed.

10 Going local

In this section I turn away from attempts to avoid the problem of belief revision in the face of zero probability events by chasing regularity, using Popper functions/full conditional probabilities, or the like. And rather than pursue a global solution I will illustrate how belief revision in the face of zero probability events may work on the local, problem specific level. Since the example comes from QM a few more technical notions from quantum probability are needed as background.

First some definitions. A projection $F_\phi \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ is a filter for a state ϕ on $\mathfrak{B}(\mathcal{H})$ just in case for any normal state ω such that $\omega(S_\phi) \neq 0$

$$\frac{\omega(S_\phi E S_\phi)}{\omega(S_\phi)} = \phi(E) \;\text{ for all } E \in \mathcal{P}(\mathfrak{B}(\mathcal{H})).$$

Among the normal states on $\mathfrak{B}(\mathcal{H})$ are the vector states. That ψ is a vector state means that there is a vector $|\psi\rangle \in \mathcal{H}$ such that $\psi(A) = \langle\psi|A|\psi\rangle$ for all $A \in \mathfrak{B}(\mathcal{H})$.[18] For a normal state the support projection is the smallest projection to which the state assigns probability 1. For a vector state ψ the support projection S_ψ is the projection onto the ray spanned by $|\psi\rangle$. Armed with these definitions it is not hard to establish that the support projection for a vector state is a filter for that state (see Ruetsche and Earman 2011).[19]

Now consider Oscar whose initial credence function Pr on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ is non-regular because it assigns zero credence to the support projection S_ψ of a vector state ψ. Suppose that Oscar makes a measurement of S_ψ which returns a Yes answer and that he is convinced that this answer is correct. How is Oscar to revise his prior credences in the light of the information that S_ψ is true? He cannot Lüders update since for him $\Pr(\bullet//S_\psi)$ is undefined. But suppose he wants his new credence function $\Pr_{new}$ to assign a probability of 1 to S_ψ and he also wants $\Pr_{new}$ to be countably additive when $\mathcal{H}$ is separable (and completely additive when $\mathcal{H}$ is non-separable). These two requirements uniquely fix Oscar's new credences over $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$. By Gleason's theorem any $\Pr_{new}$ satisfying the additivity constraint extends uniquely to a normal state ω on $\mathfrak{B}(\mathcal{H})$. So if $\Pr_{new}(S_\psi) = 1$ then $\omega(S_\psi) = 1$. Using the Cauchy-Schwartz inequality it is easy to see that this implies that $\omega(S_\psi E S_\psi) = \omega(E)$ for all $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$. So $\Pr_{new}(E) = \omega(E) = \dfrac{\omega(S_\psi E S_\psi)}{\omega(S_\psi)}$, and by the filter property of S_ψ the last expression is equal to $\psi(E)$. Note that the uniquely determined $\Pr_{new}$ is *not regular*. Regularity for real valued $\Pr_{new}$ is possible when $\mathcal{H}$ is separable since there are only a countable number of mutually orthogonal projections, but updating to a regular real valued $\Pr_{new}$ is *not* an option in the present scenario.

This unique determination of a new credence function can be generalized to cover cases of uncertain learning over some types of partitions of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$.

[18] Vector states are pure, i.e. cannot be expressed as a non-trivial convex linear combination of distinct states. And for $\mathfrak{B}(\mathcal{H})$—but not for other von Neumann algebras—the vector states are identical with the normal pure states.

[19] It follows that if ψ is a normal pure state (= vector state) on $\mathfrak{B}(\mathcal{H})$ and S_ψ is the support projection for ψ then $\Pr(E//S_\psi) = \psi(E)$ for all $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ provided that $\Pr(S_\psi) \neq 0$. This can be regarded as a version of David Lewis' Principal Principle; see Earman (2020).

Let $\{|\psi_a\rangle\} \in \mathcal{H}$ be an ON basis for $\mathcal{H}$. Suppose that Oscar's learning experience leads him to adapt new credences $\Pr_{new}(S_{\psi_a})$ on the partition $\{S_{\psi_a}\}$ of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ given by the support projections S_{ψ_a} for the vector states ψ_a corresponding to the $|\psi_a\rangle$. And suppose that Oscar wants to extend $\Pr_{new}$ from the partition $\{S_{\psi_a}\}$ to a completely additive $\overline{\Pr_{new}}$ on all of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$. There is a unique answer to his desire. By Gleason's theorem there is for any such extension a unique normal state ω on $\mathfrak{B}(\mathcal{H})$ such that $\overline{\Pr_{new}}(E) = \omega(E)$ for all $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$. Since $\sum_a S_{\psi_a} = I$, $E = E\sum_a S_{\psi_a}$ for all $E \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$, and $\omega(E) = \omega(E(\sum_a S_{\psi_a})) = \sum_a \omega(ES_{\psi_a})$. But for any a either $ES_{\psi_a} = 0$ or else $ES_{\psi_a} \neq 0$, and in the latter case $ES_{\psi_a} = S_{\psi_a}$ since the S_{ψ_a} are minimal projections. Thus, $\omega(E) = \sum_{a'} \omega(S_{\psi_{a'}})$ where the sum is now over all $S_{\psi_{a'}}$ such that $ES_{\psi_{a'}} \neq 0$. The upshot is that $\overline{\Pr_{new}}$ is uniquely determined by its values on $\{S_{\psi_a}\}$. Note that unless $\Pr_{new}(S_{\psi_a}) > 0$ for all S_{ψ_a} in the partition the new credence function $\overline{\Pr_{new}}$ is *not regular*. Again, while regularity for real valued $\overline{\Pr_{new}}$ is possible when $\mathcal{H}$ is separable, updating to a regular real valued $\overline{\Pr_{new}}$ is *not* an option in the present scenario for agents whose agnoticism does not extend to all the elements of the partition $\{S_{\psi_a}\}$.

Admittedly the cases treated here are special. But the study of problem-specific cases such as these may give clues to how a learning experience that results in new credences for a sub-family of propositions can serve to uniquely fix or otherwise constrain credences for the entire family.

11 Epilog

The literature on the problem of updating on zero-probability events focuses mainly on devices for dodging the problem. None of these devices offers a satisfactory stable resolution over repeated updatings. Furthermore, the literature is devoted almost exclusively to classical probability, and there is no serious attempt show how the devices developed for classical probability can be deployed for quantum probabilities.

I suggest that it is high time to confront the problem rather than trying to dodge it. Stick to real valued probability measures, and forget about infinitesimal probabilities, Popper functions/full conditional probabilities, and other nostrums which only postpone the problem. Come to terms with the fact that for real valued probability functions non-regularity must be countenanced: when the domain of the function is sufficiently 'fat' regularity may not be possible, and even when the domain is 'slim' enough for regularity to be possible recognize that regularity is not always a desirable regulatory ideal nor is it sustainable across learning experiences without using artificial contortions.

In the context of classical probability the problem to be solved is this: Suppose an agent's initial credences are represented by a standard classical probability space $(\Omega, \Sigma, pr(\bullet))$ where $pr(\bullet)$ is not regular. The agent has a learning experience resulting in new credences $pr_{new}(F_a)$ for the elements $F_a \in \widehat{\Sigma}$ of a proper subset $\widehat{\Sigma}$ of Σ, where $pr(F_{a^*}) = 0$ for one or more of the F_a's whereas $pr_{new}(F_{a^*}) > 0$. The challenge: (1) supply an extension of the pr_{new} for $\widehat{\Sigma}$ to a probability measure on all of Σ, and either (2) prove the extension is unique or, failing (2), (3) justify the chosen extension as providing the rational way to revise the initial pr-credences for Σ in light of the newly acquired credences pr_{new} for $\widehat{\Sigma}$. For quantum probabilities the challenge is analogous with $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ in place of Σ and a proper subset of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ in place of $\widehat{\Sigma}$.

In the preceding section it was seen that there are cases in QM where either gaining probabilistic certainty ($\mathrm{Pr}_{new}(F) = 1$) for a judiciously chosen element $F \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ or simply shifting to new uncertainties ($0 \leq \mathrm{Pr}_{new}(F_a) \leq 1$) for a judiciously chosen partition $\{F_a\} \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ suffices to fix a unique extension *regardless of the initial* Pr. While interesting and instructive such cases are hardly typical. More typical are cases where the newly acquired credences by themselves do not suffice to fix a unique extension. What then are Bayesians to do?

One response would be to identify and justify additional rationality constraints, e.g. minimal change principles, to constrain the possible extensions. Another would be to eschew a unified response in favor of tailoring solutions to fit context-dependent goals of inquiry. More radically, Bayesisian personalists can have the courage of their (non-) convictions. What I mean to indicate here is the attitude that to the extent that the newly acquired credences, together with the initial credences and uncontroversial rationality principles, fail to constrain extensions then each Bayesian agent is free to choose her own way of revising her initial probability measure in light of her newly acquired credences, just as she was free to choose her initial credences. In line with this sentiment, some commentators who reject the diachronic Dutch book justifications for Bayes and for Jeffrey conditionalization are led to the conclusion that, even when zero probability events are not at issue, updating in light of a learning experience is up to the individual agent unless the agent's newly acquired credences fix a unique extension. While such libertarianism about updating is the logical outcome of a thoroughgoing personalist interpretation of probability it would seem to undermine any hope that an island of objectivity might rise from the sea of subjective opinion in the form of merger of opinion among the members of the community of fellow Bayesian agents as a result of repeated updatings on accumulated evidence. While thoroughgoing personalists may not take such loss of hope as a criticism, it will be disappointing to those who want to use Bayesianism to explain and defend the objectivity of science.[20]

I do not presume to tell Bayesians which of these responses—or some other entirely different response—to pursue. But I insist that an honest and worthy Bayesianism should confront the challenge of updating on zero probability events rather than trying to avoid it and, further, I insist that a worthy Bayesianism confront the challenge for quantum probabilities as well as classical probabilities. This is not a tale of the emperor's new clothes; but it is a tale where the Emperor Bayes is rather scantily clad, and the fan dance of infinitesimal probabilities, Popper functions, and other nostrums does not succeed in hiding the scantiness.

[20] The degree-of-belief interpretation of probability is exactly what is called for in the context of personal decision making under uncertainty. But unless the personal probabilities of Bayesian agents bear some systematic relation to truth, Bayesianism cannot underwrite the objectivity of scientific inference. The most widely discussed response to this challenge is to appeal to results about convergence to the truth in the form of convergence to probability 1 under repeated updating on accumulating evidence. Obviously, however, updating by Bayes/Lüders conditionalization will not produce such convergence if the true hypothesis has zero prior probability. A rule for updating on zero probability events would potentially allow the truth to emerge—but not if the "rule" is freelance libertarianism.

Appendix

Let (Ω, Σ, pr) be a standard classical probability space. That pr is finitely (respectively, countably, completely) additive means that for any finite (respectively, countable, uncountable) collection of pairwise disjoint $E_a \in \Sigma$, $pr(\cup_a E_a) = \sum_a pr(E_a)$. When the sum $\sum_a pr(E_a)$ is over an uncountable index set $\mathcal{I}$ it is understood as $\lim_F \sum_a pr(E_a)$ where the F are finite subsets of $\mathcal{I}$, and $\lim_F \sum_a pr(E_a) = L$ means that for any $\epsilon > 0$ there is a finite $F_0 \subset \mathcal{I}$ such that for any finite F with $\mathcal{I} \supset F \supset F_0$, $|\sum_a pr(E_a) - L| < \epsilon$.

Let $\{E_a\} \in \Sigma$ be a family of pairwise disjoint sets. Then at most a countable number of the E_a get positive probability. To see this define $S_n =:$ $\{E_a : \dfrac{1}{n+1} \leq pr(E_a) \leq \dfrac{1}{n}\}$, $n = 1, 2, 3, \ldots$. By finite additivity each of the S_n has only a finite number of members. The union $\cup_{n=1}^{\infty} S_n$ contains all the elements of the family that have positive probability. But the countable union of finite sets is countable.

Let κ be the least measurable cardinal. This means that it is the least cardinal such that there is a set Ω, where $|\Omega| = \kappa$, and a probability space $(pr, \Omega, P(\Omega))$, where $P(\Omega)$ is the power set of Ω, such that $pr(\{x\}) = 0$ for every $x \in \Omega$. Measurable cardinals are gigantic—they lie beyond the hierarchy of infinities $\aleph_0, \aleph_1, \aleph_2, \ldots$ that Cantor bequeathed us.

> *Lemma*: If $\alpha < \kappa$ then every countably additive measure on $P(\Omega)$, $|\Omega| = \alpha$, is completely additive.

The proof of the Lemma is by reductio. Suppose that pr is a countably additive measure on $P(\Omega)$, where the set Ω has cardinality $\alpha < \kappa$, and that pr fails to be completely additive. Then there is an uncountable collection of pairwise disjoint $E_a \in P(\Omega)$ of cardinality less than or equal to α such that

$$(\dagger) \quad pr(\cup_a E_a) \neq \textstyle\sum_a pr(E_a).$$

Case (a) $pr(E_a) = 0$ for all the E_a, which implies by $(\dagger)$ that $pr(\cup_a E_a) = r > 0$. Consider the set $\widetilde{\Omega}$ whose elements are the E_a. Define a measure $\widetilde{pr}$ on $P(\widetilde{\Omega})$ by $\widetilde{pr}(Y) := \dfrac{pr(\cup_{a \in Y} E_a)}{r}$. This is a countably additive measure, contradicting the assumption that κ is the least measurable cardinal since $|\widetilde{\Omega}| \leq \alpha$ and $\widetilde{pr}(\{E_a\}) = 0$ for all $E_a \in \widetilde{\Omega}$.

Case (b) $pr(E_a) > 0$ for some E_a. There can be only a countable number of E_a with positive measure. Using the fact that pr is countably additive these E_a can be removed so that without loss of generality this case is reduced to Case (a). (In more detail: Denote the elements E_a that have non-zero measure by E_n, $n \in \mathbb{N}$, and define $T := \cup_n E_n$. By countable additivity $pr(T) = \sum_n pr(E_n) = s > 0$, and, by (†), $s < 1$. Define a new countably additive measure on $P(\Omega)$ by $\widehat{pr}(\bullet) := \dfrac{pr(\bullet \cap \overline{T})}{1-s}$. This measure falls under Case (a).)

An immediate consequence of the Lemma is that Lebesgue measure on $[0, 1] \subset \mathbb{R}$ cannot be extended to a countably additive measure on $P([0, 1])$ if the cardinality of the continuum is less than the least measurable cardinal.

Somewhat less obvious is an implication for quantum probability:

> *Cor.* Let Pr be a countably additive quantum probability measure on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$. Unless $\dim(\mathcal{H})$ is as great as the least measurable cardinal Pr is completely additive.

The idea of the proof is simple. Choose a basis B for $\mathcal{H}$. The subsets of B are in one-one correspondence with the closed subspaces of $\mathcal{H}$ which in turn are in one-one correspondence elements of $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$. Thus, corresponding to a quantum probability Pr on $\mathcal{P}(\mathfrak{B}(\mathcal{H}))$ there is a classical probability space $(B, P(B), pr)$ where $P(B)$ is the power set of B and where $pr(b)$, $b \in P(B)$, is equal to $\Pr(E_b)$ where $E_b \in \mathcal{P}(\mathfrak{B}(\mathcal{H}))$ is the projection corresponding to b. The Lemma can now be applied. For details see Drish (1979) and Eilers and Horst (1975).

References

[1] Armendt, B. 1980. "Is There a Dutch Book Argument for Probability Kinematics?" *Philosophy of Science* 47: 583-588.

[2] Bernstein, A. R. and Wattenberg, F. 1969. "Nonstandard Measure Theory," in A. J. Luxenburg (ed.), *Applications of Model Theory to Algebra, Analysis, and Probability*, pp. 171-185. New York: Holt, Rinehart and Winston.

[3] Bub, J. 1977. "Von Neumann's Projection Postulate as a Possible Conditonalization Rule in Quantum Mechanics," *Journal of Philosophical Logic* 6: 381-390.

[4] Cassinelli, G. and Zanghi, N. 1983. "Conditional Probabilities in Quantum Mechanics," *Nuovo Cimento B* 73: 237-245.

[5] Christensen, D. 1991. "Clever Bookies and Coherent Beliefs," *Philosophical Review* 100: 229-247.

[6] Drish, T. (1979). "Generalizations of Gleason's Theorem," *International Journal of Theoretical Physics* 18: 239-243.

[7] Dubins, L. 1975. "Finitely Additive Conditional Probabilities, Conglomerability and Disintegrations," *Annals of Probability* 3: 89-99.

[8] Earman, J. 2018. "The Relation between Credence and Chance: Lewis' 'Principal Principle' Is a Theorem of Quantum Probability Theory." http://philsci-archive.pitt.edu/id/eprint/14822

[9] _______ 2019a. "Quantum Bayesianism Assessed." *Monist* 102: 403–423.

[10] _______ 2019b. "Lüders conditionalization: Conditional probability, transition probability, and updating in quantum probability theory," http://philsci-archive.pitt.edu/id/eprint/15914.

[11] _______ 2020. "Implementing David Lewis' Principal Principle: A Program for Investigating the Relation between Credence and Chance," http://philsci-archive.pitt.edu/id/eprint/18201.

[12] Eilers, M. and Horst, E. (1975). "Theorem of Gleason for Nonseparable Hilbert Spaces," *International Journal of Theoretical Physics* 13: 419-424.

[13] Gallow, J. G. 2019. "Diachronic Dutch Books and Evidential Import," *Philosophy and Phenomenological Research* XCIX: 49-80.

[14] Halpern, J. Y. 2009. "Lexicographic probability, conditional probability, and nonstandard probability," arXiv: cs/0306106.

[15] Hamhalter, J. 2003. *Quantum Measure Theory*. Dordrecht: Kluwer Academic.

[16] Hofweber, T. 2014. "Cardinality Arguments Against Regular Probability Measures," *Thought* ISSN 2161-2234.

[17] Hofweber, T. and Schindler, R. 2016. "Hyperreal-Valued Probability Measures Approximating a Real-Valued Measure," *Notre Dame Journal of Formal Logic* 57: 369-374.

[18] Israel, J. 2008. "*Raid!* Dissolving the Big, Bad Bug," *Noûs* 42: 291-307.

[19] Jackson, F. 1987. *Conditionals*. Oxford: Blackwell.

[20] Jeffrey, R. C. 1965. *The Logic of Decision*. Chicago: University of Chicago Press.

[21] ______ , 1992. *Probability and the Art of Judgment*. Cambridge: Cambridge University Press.

[22] Kadison, R. V. and Ringrose, J. R. 1991. *Fundamentals of the Theory of Operator Algebras*, 2 Vols. Providence, RI; American Mathematical Society.

[23] Kemeny, 1955. "Fair Bets and Inductive Probabilities," *Journal of Symbolic Logic* 20: 263–273.

[24] Krauss, P. H. 1968. "Representation of Conditional Probability Measures on Boolean Algebras," *Acta Mathematica Academiae Scientiarum Hungaricae Tomas* 19: 229-241.

[25] Lewis, D. L. 1980. "A Subjectivist Guide to Objective Chance." In R. Carnap and R. Jeffrey (eds.), *Studies in Inductive Logic and Probability*, pp. 263-293. Berkeley: University of California Press.

[26] McGee, V. 1994. "Learning the impossible," in E. Ellis and B. Skyrms (eds.), *Probability and Conditionals*, pp. 179-199. Cambridge: Cambridge University Press.

[27] Popper, K. R. 1961. *The Logic of Scientific Discovery.* New York: Science Editions.

[28] Pruss, A. 2013. "Probability, Regularity, and Cardinality," *Philosophy of Science* 80: 231-240.

[29] Ruetsche, L. 2011. "Why Be Normal?" *Studies in History and Philosophy of Modern Physics* 42: 107-115.

[30] Ruetsche, L. and Earman, J. 2011. "Interpreting probabilities in quantum field theory and quantum statistical mechanics," in C. Beisbart and S. Hartmann (eds.), *Probabilities in Physics,* pp. 263-290. Oxford: Oxford University Press.

[31] Shimony, A. 1955. "Coherence and the Axioms of Confirmation," *Journal of Symbolic Logic* 20: 1-28.

[32] Skyrms, B. 1980. *Causal Necessity.* New Haven, CN: Yale University Press.

[33] _________ 1987. "Dynamics Coherence and Probability Kinematics," *Philosophy of Science* 54: 1-20.

[34] Teller, Paul, 1976, "Conditionalization, Observation, and Change of Preference", in W. Harper and C. A. Hooker (eds.), *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science.* Dordrecht: D. Reidel.

[35] Vineberg, S. 2016. "Dutch Book Arguments," *Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/entries/dutch-book/.

[36] von Baeyer, H. C. 2016. *QBism: The Future of Quantum Physics.* Cambridge. MA: Harvard University Press.

[37] Weisberg, J. 2011. "Varieties of Bayesianism," in D. Gabbay, S. Hartmann, and J. Woods (eds.), *Handbook for the History of Logic Vol. 10, Inductive Logic*. Amsterdam: Elsevier.