Revising Probabilities and Full Beliefs

Sven Ove Hansson¹

Received: 15 June 2019 / Accepted: 6 January 2020 / Published online: 19 March 2020
© The Author(s) 2020

Abstract

A new formal model of belief dynamics is proposed, in which the epistemic agent has both probabilistic beliefs and full beliefs. The agent has full belief in a proposition if and only if she considers the probability that it is false to be so close to zero that she chooses to disregard that probability. She treats such a proposition as having the probability 1, but, importantly, she is still willing and able to revise that probability assignment if she receives information that gives her sufficient reasons to do so. Such a proposition is (presently) undoubted, but not undoubtable (incorrigible). In the formal model it is assigned a probability $1 - \delta$, where δ is an infinitesimal number. The proposed model employs probabilistic belief states that contain several underlying probability functions representing alternative probabilistic states of the world. Furthermore, a distinction is made between update and revision, in the same way as in the literature on (dichotomous) belief change. The formal properties of the model are investigated, including properties relevant for learning from experience. The set of propositions whose probabilities are infinitesimally close to 1 forms a (logically closed) belief set. Operations that change the probabilistic belief state give rise to changes in this belief set, which have much in common with traditional operations of belief change.

Keywords Probability dynamics · Update · Probability revision · Multistate model · Probabilistic learning · Belief change · Infinitesimal probabilities

1 Introduction

Formal models of belief states can employ either a dichotomous (all-or-nothing) or a more fine-graded representation of beliefs. The former approach is applied in models belonging to the tradition called belief change or belief revision [1, 5, 9, 15, 18, 35]. In these models, an agent's current beliefs are represented by a consistent and logically closed set of propositions, called a "belief set". We will call these models

---

✉ Sven Ove Hansson
soh@kth.se

¹ Division of Philosophy, Royal Institute of Technology (KTH), Teknikringen 76, 100 44 Stockholm, Sweden

dichotomous since they have only two degrees of belief; a proposition is either believed, in which case it is an element of the belief set, or it is not believed, and not an element of the belief set. The most important fine-grained approach is probability theory, in which a belief state is represented by a probability function with a dense range of infinitely many degrees of belief, between 0 and 1.

Both dichotomous and probabilistic representations are usually treated as parts of dynamic frameworks, which also contain operations of change. Such operations specify how the belief state will be modified in response to various inputs. In both approaches, the most important type of input is a proposition to be assimilated as a full belief. If the input is a proposition a , then in a dichotomous belief change model, a will be an element of the new belief set that results from the operation. In a probabilistic model, the new probability function will assign the probability 1 to a .

In standard probability theory, the new probability function p' is commonly assumed to be completely determined, through conditionalization, by the input a and the previous probability function p :¹

$$p'(d) = p(d \mid a) = \frac{p(a \& d)}{p(a)} \quad (1)$$

In contrast, the previous belief set K in a dichotomous belief change model does not contain enough information for determining, in combination with the input a , the contents of the resulting new belief set. The major reason for this is that if a is inconsistent with K , then some elements of K will have to be removed in order to retain consistency when a is added. There are many ways to do this, and K alone does not tell us how to do it. Therefore, the belief state has to contain more information than what can be inferred from the belief set K ; this additional information should tell us which elements to exclude from K in order to make room for a . Quite a few types of such additional components of the belief state (“selection mechanisms”) have been proposed (Rott and Hansson [36]; Hansson [18], pp. 6–16). Some of them take the form of relations ordering the elements of K , or numbers assigned to its elements to signal their degrees of retractability. However, such numerical assignments do not denote degrees of beliefs, and they do not have the properties expected of plausible such degrees [9, pp. 87–88].

Another important difference between dichotomous and probabilistic models of belief change concerns the removability of full beliefs. In the common models of (dichotomous) belief change, a non-tautologous belief that is an element of the belief set can be removed in a subsequent operation of change. For instance, with $*$ denoting the operation of change and a a non-tautologous input proposition, we have $a \in K * a$ and $a \notin K * a * \neg a$, according to the standard properties of such operations.² In contrast, as can be seen from Eq. 1, conditionalization by a proposition a that has

¹ Unless $p(a) = 0$, in which case $p'(d)$ is left undefined by Eq. 1. This equation is offered as a definition in Kolmogorov [27, p. 6], but as noted by Hájek [10, p. 274] it should preferably be called an analysis (“the ratio analysis”), rather than a definition. Kolmogorov also offered an account of conditional probability in terms of Markov chains (ibid., pp. 12–13), which allows conditional probabilities with a zero-probability antecedent to be well-defined. The problems with the ratio analysis are nicely summarized in Lyon [33]. Unfortunately, Kolmogorov’s account in terms of Markov chains has its own problems (Seidenfeld et al. [39]; Seidenfeld [38]; Hájek [10], pp. 291, 301, and 319).

² $a \notin K * a * \neg a$ follows from the AGM postulates success and consistency [1].

non-zero probability results in a new probability function p' with $p'(a) = 1$, and any series of further conditionalizations of p' will result in a probability function that also assigns the value 1 to a .

Probabilistic and dichotomous belief representations capture different features of real-life epistemic change. Probabilistic models have the obvious advantage of allowing for smaller changes than a shift between full belief and complete lack of belief. On the other hand, the common probabilistic models are not well suited to represent the process of giving up a full belief. The gist of the problem is that we all have beliefs which we do not doubt, but which we will nevertheless give up if confronted with sufficiently convincing evidence that they are mistaken. Such a pattern is difficult to represent in probabilistic models, since they use the same means (probability 1) to represent that a belief is not doubted and that it cannot be rejected.³ The loss of full beliefs is much more readily represented in a dichotomous belief change model. Indeed, operations that remove beliefs from the belief set (operations of contraction) are standard components of such models.

Several attempts have been made to reconcile these two types of models with each other. (For overviews, see Leitgeb [31], pp. 1-53 and Zhuang et al. [47].) In most of these proposals, an individual's probabilistic and all-or-nothing beliefs refer to the same propositions, so that each proposition has both a probability and a dichotomous belief status (believed/not believed). The two assignments of a belief status are usually closely connected to each other in these models; typically a proposition's dichotomous belief status can be inferred from the probability function. The most obvious such approach is to identify the set of full beliefs with the propositions that have probability 1. However, in combination with standard conditionalization, this construction has the rather devastating effect that once a full belief has been acquired, there is no way to give it up, whatever new information the agent receives.⁴ One

³To avoid this, we need a framework in which only undoubtable propositions have unit probability, or equivalently: only propositions with undoubtable negations have zero probability. This is close to requiring *regularity*, a property promoted most famously by David Lewis, who defined it as meaning that only "the empty proposition, true at no worlds" has zero probability [32, p. 88]. He noted that if the sample space is infinite, then regularity implies that many of its elements have infinitesimal probabilities. Brian Skyrms pointed out that in such a sample space, failure of regularity is accompanied by failure of Shimony's [40] strict coherence property; betting on the truth of a contingent proposition with probability 1 will be "a bet which we will consider fair even though we can possibly lose it but cannot possibly win it." [41, p. 74]

Importantly, regularity fails in standard probability theory even if the sample space is finite. Common Bayesian revision of a probability function p by a contingent proposition e results in a new probability function p' with $p'(a) = p(a \mid e)$ for all a , and consequently $p'(\neg e) = 0$ even though $\neg e$ is contingent. With a revision function such as that introduced below in Definition 5, this will not be the case.

⁴This leads to a major limitation in standard probabilistic models, namely that they cannot represent revision by a proposition previously believed to be false. In dichotomous belief change, such an operation is represented by a revision $K * a$ for an a with $\neg a \in K$. The operations of revision studied in belief change theory determine what $K * a$ will be with the help of some mechanism that selects those parts of K that will be retained in $K * a$. (Typically, the retained parts of K will not imply $\neg a$ unless it is a tautology, and therefore a can be added without loss of consistency.) In standard probabilistic representation, revision by a sentence a corresponds to conditionalization by a . The outcome is a new probability function p' that replaces the original probability function p , such that $p'(d) = p(d \mid a)$ for all d . However, if $p(\neg a) = 1$, then $p(d \mid a)$ is undefined according to standard accounts of conditional probability. (See footnote 1.)

way to solve this would be to apply the so-called Lockean thesis [6], which identifies the full beliefs with the propositions that have a probability higher than some fixed probability limit below 1. However, no non-arbitrary such limit is within sight, and in addition, this construction is incompatible with the common assumption that what follows logically from two beliefs is also believed.⁵ Recently, other more sophisticated constructions have gained interest, in which the criterion for status as full belief is not a fixed minimal level of probability but some comparison with the probabilities of propositions representing alternative beliefs on the same subject-matter. Such constructions are highly sensitive to the individuation of said alternatives. In Leitgeb's [31] proposal, which is arguably the best worked-out of these constructions, full belief will sometimes be assigned to propositions with a probability barely above 0.5, but on other occasions it will not be assigned to propositions with very high probabilities [19, pp. 278-279].⁶

In this article, a new approach will be presented. Its basic idea is that we have full belief in a proposition if and only if we consider the probability that it is false to be so close to zero that we choose to disregard that probability.⁷ We therefore treat the proposition as if it had the probability 1, but, importantly, we are still willing and able to revise that probability assignment if we receive information that gives us sufficient reasons to do so. Another way to express this is that such a proposition is (presently) undoubted, but not undoubtable (in corrigible). As I have argued elsewhere, this is the common form of full beliefs, both in science and in everyday life [13], [20, pp. 68-71].

In order to represent undoubted but doubttable propositions, we need to enrich the standard representation of probabilities with some means to keep track of probabilities that are small enough to be currently neglected, but still capable of being raised to levels at which the propositions they refer to will be taken seriously. For this purpose,

⁵If I believe that Manchester United will win the match they play on Sunday in London, and I also believe that Flamengo will win the match they are playing in Rio de Janeiro on the same day, then I will expectedly also believe that both teams are going to win their matches on that day. However, if the first two predictions are just above the probability limit for full belief, then the third, conjunctive statement can fall below it. It is commonly assumed that the set of beliefs held by a rational agent should be closed under logical consequence. Therefore, the observation that what follows from two propositions that are both above a certain probability limit can fall below that limit is commonly taken as a decisive argument against using a real-valued probability limit below 1 as criterion of rational belief. This is highlighted in two of the most famous probability paradoxes, namely the lottery paradox [30, pp. 197-198] and the paradox of the preface [34].

⁶See Schurz [37] for a series of impossibility results showing that the combination of the Lockean thesis (with a real-valued probability limit below 1) and logical closure of the set of beliefs can only obtain in models that are small in terms of the number of beliefs and/or the number of doxastic possibilities that are represented.

⁷The probabilities referred to in this article are "objectivist credences", i.e. they will represent the agent's best estimate of objective probabilities.

we will use hyperreal numbers to represent probabilities.⁸ A (finite) hyperreal number is either a real number or the sum of a real number and a (positive or a negative) infinitesimal number. (A positive infinitesimal is larger than 0 but smaller than all positive real numbers. See Appendix A for a brief introduction to hyperreal numbers.) The following notation will be used:

Definition 1 (1) Propositions are denoted by the letters a, b, c, d , and e .

(2) The letters s, t, u, v, x, y , and z represent hyperreal numbers (which may be real). The letter δ represents numbers that are either 0 or infinitesimal, and ϵ represents infinitesimal numbers.

(3) The standard (real) part of a finite hyperreal number s is denoted $\text{st}(s)$.

(4) The symbols $\approx$, $\not\approx$, $\lesssim$, and $\ll$ are used as follows:

$a \approx b$ if and only if $\text{st}(a) = \text{st}(b)$

$a \not\approx b$ if and only if $\text{st}(a) \neq \text{st}(b)$

$a \lesssim b$ if and only if $a < b$ and $\text{st}(a) = \text{st}(b)$

$a \ll b$ if and only if $\text{st}(a) < \text{st}(b)$

Probabilities will be represented by hyperreal numbers in the closed interval $[0, 1]$. The agent has full belief in a proposition a if and only if $\text{st}(\text{p}(a)) = 1$.⁹ If a is a

⁸There is a considerable literature on infinitesimal probabilities. Most of it has employed infinitesimals to attack the problems arising when classical probability theory is applied to infinite domains (event spaces). In the words of Brian Skyrms, infinitesimal probabilities are needed since we have “fattened the domain of the probability function (the space of possibilities) without a concomitant fattening of the range” [41, p. 74]. For instance, a fair lottery with an infinite number of tickets cannot be modelled with real-valued probabilities, but it can be modeled by assigning the same, non-zero, infinitesimal probability to all tickets [2, 44, 45]. For a recent overview, see Benci et al. [3]. In this article, the domain of the probability function will be assumed to be finite, and therefore we will not need infinitesimal probabilities for their most common purpose, namely to ensure that all logically possible outcomes receive non-zero probability. Instead, we will use infinitesimals for two other purposes, both of which are known from previous literature: (1) we will identify the set of full beliefs with the set of propositions whose probability is infinitesimally close to 1 [44], and (2) we will use propositions with infinite probabilities as “memory tracks” of beliefs that have been given up [43]. Major new features in this article are the transfer of the distinction between update and revision from the belief change literature (Section 2.2), the combination of infinitesimals with an approach to probability revision that differs from traditional Bayesian conditionalization (Sections 2.3–2.4), and several adjustments that align revisions of probabilities with the operations on belief sets that are studied in the belief change literature.

⁹This can be described as a non-standard version of the Lockean thesis. Essentially the same version of the thesis was applied by Wenmacker [44] in a study of fair lotteries with an infinite number of tickets. (However, she employed Hrbacek’s [23] relative analysis, which mirrors infinitesimal numbers as a category of ‘ultrasmall’ real numbers.)

Contrary to a limit for full belief at (arbitrarily small) positive real-valued distance from 1, a limit infinitesimally smaller than 1 is not affected by Schurz’s [37] impossibility theorems. To see why this is so, consider his result for the sufficiency part of the Lockean thesis. According to this part of the thesis, there is a real-valued number t with $1/2 < t \leq 1$, such that a proposition a is a full belief if $t \leq \text{p}(a)$. Suppose that there are n logically independent full beliefs $a_1, \dots, a_n$ such that for all a_i : $t \leq \text{p}(a_i) < u < 1$, where u is an upper bound on their probabilities. Due to the logical closure of the set of full beliefs, $t \leq \text{p}(a_1 \& \dots \& a_n)$. Due to the probability axioms, $\text{p}(a_1 \& \dots \& a_n) = \text{p}(a_1) \times \dots \times \text{p}(a_n) < u^n$. Thus $1/2 < u^n$. For instance, if $u = 0.95$, then $n < 14$ and if $u = 0.98$, then $n < 35$. (Schurz presents this result in a generalized form that does not presume logical independence.)

However, in the hyperreal model, there is no justification for imposing an upper bound such as u . Even if we do so, it follows from $t \lesssim u$ that $1/2 < u^n$ for all n .

tautology, then $p(a) = 1$, but we will leave it open whether there are non-tautologies with unit probability.

In this article, it will be assumed that the domain (sample space) of the probability function is finite. Therefore, we can also assume that the standard axioms for probability, the Kolmogorov axioms, hold in spite of the codomain (range) of the probability function being extended from the closed real-valued interval $[0, 1]$ to the closed hyperreal-valued interval $[0, 1]$.¹⁰ Probabilistic revision will be performed with an operation of revision that does not in general coincide with standard (Bayesian) conditionalization.

Importantly, no metaphysical claims are made about infinitesimals. They are used here as a modelling tool since they have properties that are convenient for the representation of probabilities that are conceived as so small that they need not be taken into account. Other constructions, such as vectors of real numbers, could be used for the same purpose, but the hyperreal number system has the advantage of being well established and having well-known properties.

The new model of probabilistic revision is introduced in Section 2. In Section 3 some of its properties are investigated. The set of propositions whose probability is infinitesimally close to 1 forms a (logically closed) belief set. Section 4 is devoted to the changes in that belief set that follow from operations of probabilistic change. All formal proofs are deferred to an appendix.

2 Constructing the Multistate Model of Probabilistic Revision

In order to construct a workable model of epistemic change, based on the principles introduced in the previous section, some additional adjustments of the traditional account of probabilistic revision are needed. These changes will bring the probabilistic model closer to the common models of dichotomous belief change. To begin with, we will adopt an iteration-friendly notation for revision from the dichotomous models. After that, we will transfer the distinction between update and revision from dichotomous modelling. With these adjustments in place, we are ready to introduce a new model of probabilistic belief states and their revision, to be called the multistate model.

2.1 An Iteration-Friendly Notation

The common way to express probabilistic change in formal language is to employ the notation for conditionalization. In standard (real-valued) probability theory, the outcome of updating a probability function p by a proposition a_1 to which p assigns a non-zero probability is a new probability function p' , which is usually written

¹⁰If the sample space is infinite, then the axiom of countable additivity,

$$p\left(\bigcup_{n=1}^{\infty} e_n\right) = \sum_{n=1}^{\infty} p(e_n)$$

may not hold since the sum referred to on the right-hand side of the equation may not exist. (Benci et al 2013, p. 125)

$p(\cdot \mid a_1)$, i.e. $p'(d) = p(d \mid a_1)$ for all d . This notation is impractical for repeated operations of change. For instance, if p' is in its turn revised by the proposition a_2 , then this results in a new probability function p'' such that $p''(d) = p'(d \mid a_2)$ for all d , and in this notation the preceding revision by a_1 is lost out of sight. The common notation in the literature on dichotomous belief change is much more clear in this respect. The revision of a belief set K by a proposition a_1 is denoted $K * a_1$, its revision by first a_1 and then a_2 is denoted $K * a_1 * a_2$, etc. This notation can be used for probabilistic revision as well. Thus, the revision of a probability function p by a proposition a_1 will be written $p * a_1$, its revision by first a_1 and then a_2 will be denoted $p * a_1 * a_2$, etc. Hence, $p * a_1 * a_2(d) = 0.7$ means that d is assigned the probability 0.7 by the probability function that is obtained by revising p by first a_1 and then a_2 . All inputs for the revision operation $*$ will be propositions, and $*a_1$ represents the receipt of information indicating that a_1 is the case.

Formulas such as $p * a_1 * a_2(d)$ are somewhat difficult to parse. For clarity, we will introduce boldface brackets around subformulas denoting a composite probability function, thus writing $(p * a_1 * a_2)(d)$ instead of $p * a_1 * a_2(d)$.

2.2 Update and Revision

In a seminal article, Keller and Wilkins [26] observed that there are two types of reasons why we incorporate new information into a belief set. One is that the world has changed, and the other that we have received new information about it. They called the first type of belief change “change-recording” and the second “knowledge-adding”. In the subsequent literature, another terminology has been established for this distinction. Belief changes in response to information about changes in the world are called “updates”, whereas belief changes based on additional information about unchanged features of the world are called “revisions”. Examples of the following type are commonly used to illustrate the difference between update and revision:

Initially, the epistemic agent believes that Mr. Booker has placed either a comic book or a dictionary, but not both, on a certain, previously empty, table.

Case i: The agent receives the additional information that it was in fact a comic book that Mr. Booker put on the table. She now believes that there is a comic book on the table, but no dictionary. This is a revision. Letting c represent the proposition that there is a comic book on the table, the revision is denoted $K * c$.

Case ii: The agent receives the additional information that someone just put a comic book on the table. She now believes that there are either two comic books on the table, or a comic book and a dictionary. This is an update, denoted $K \diamond c$.

To see that the two operations differ, let d denote that there is a dictionary on the table, and note that $\neg d \in K * c$ and $\neg d \notin K \diamond c$.

The following example shows how the distinction between update and revision can be applied to probabilistic belief states:

An urn is placed on the table. We know that it contains 100 thoroughly mixed balls that all have equal size and weight. We also know that there are either 75 yellow

and 25 black balls, or 75 black and 25 yellow ones, but we do not know which. We consider the two possible colour compositions to be equally probable.

The urn is shaken, and a ball is pulled out. It turns out to be yellow.

Problem 1: What is now the probability that a yellow ball has been drawn?

Problem 2: What is now the best estimate of what the probability was, just before the draw, that a yellow ball would be drawn?

In Problem 1, which concerns update, the probability asked for is 1. In formal terms: $(p \diamond a)(a) = p(a \mid a) = 1$. Problem 2 concerns a revision. To solve it, we can assume that there are two possible states of the world, namely b_1 in which the urn contains 75 yellow and 25 black balls, and b_2 in which it contains 75 black and 25 yellow balls. Let a be the proposition that the drawn ball is yellow. We have $p(b_1) = p(b_2) = 0.5$, $p(a \mid b_1) = 0.75$, and $p(a \mid b_2) = 0.25$. It follows directly that

$$\begin{aligned} p(b_1 \mid a) &= \frac{p(b_1 \& a)}{p(a)} = \frac{p(b_1 \& a)}{p(b_1 \& a) + p(b_2 \& a)} = \\ &= \frac{p(b_1) \times p(a \mid b_1)}{p(b_1) \times p(a \mid b_1) + p(b_2) \times p(a \mid b_2)} = 0.75 \end{aligned}$$

Thus, given our observation of a , the best estimate of the probability of b_1 is 0.75, thus $(p * a)(b_1) = 0.75$. By a similar reasoning, $(p * a)(b_2) = 0.25$. Our beliefs about the contents of the urn in each of the two states of the world have not changed, thus:

$$(p * a)(a \mid b_1) = p(a \mid b_1) = 0.75$$

and

$$(p * a)(a \mid b_2) = p(a \mid b_2) = 0.25$$

and consequently:

$$\begin{aligned} (p * a)(a) &= (p * a)(a \& b_1) + (p * a)(a \& b_2) \\ &= (p * a)(b_1) \times (p * a)(a \mid b_1) + (p * a)(b_2) \times (p * a)(a \mid b_2) \\ &= 0.75 \times 0.75 + 0.25 \times 0.25 = 0.625 \end{aligned}$$

Thus, $(p * a)(a) = 0.625$.

2.3 A Multistate Model for Real-Valued Probabilities

As the urn example shows, update corresponds to the standard account of how we “revise” (conditionalize) beliefs. Revision (in the sense used here) is of course fully in line with how problems like the above urn problem are standardly solved, but it has not been recognized as a separate operation of change. That we will do here. For that purpose, we need to collect the information about the alternative probabilistic states of the world (in the above examples, b_1 and b_2) into a single state, which will be the object to which the revision is applied. For clarity, the resulting model will first be constructed for real-valued probabilities, and then (in Section 2.4) generalized to accommodate hyperreal probabilities.

We will assume that the epistemic agent entertains an exhaustive set B of mutually exclusive possible states of the world, which are not directly observable. In formal terms:

Definition 2 A *state catalogue* for a probability function p is a non-empty set B of propositions within the domain of p , such that:

- (1) If $b \in B$, then $p(b) \neq 0$.
- (2) If $b_1, b_2 \in B$ and $b_1 \neq b_2$, then $p(b_1 \& b_2) = 0$.
- (3) $\sum_{b \in B} p(b) = 1$

Notably, the state catalogue B is only a proper subset of the domain (universe, sample space) of p . Its elements are states of the world, but not in the usual, deterministic sense. Instead, each of its elements “assigns” probabilities to propositions concerning what can be observed in the world (in short: observables). These “assignments” will take the form of observation-independent probabilities conditional on each element of B .

We will assume for simplicity that B is finite. This should be a fairly unproblematic assumption, since B does not denote the set of possible states of the world, but rather a set of (probabilistic) states that are available for consideration by the epistemic agent. It may but need not be the case that $b_1 \vee \dots \vee b_n$ is a tautology.

The empirical propositions to which each element of B “assigns” probabilities form a set E , another subset of the domain of p , which has the following properties:

Definition 3 A *set of observables* for a probability function p is a non-empty set E of propositions within its domain, such that

- (1) If $a_1, a_2 \in E$, then $a_1 \& a_2 \in E$.
- (2) If $a \in E$, then $\neg a \in E$.

Furthermore, E is *logically disjoint* from a state catalogue B if and only if no logically contingent truth-functional combination of elements of E follows logically from some logically contingent truth-functional combination of elements of B .

With the standard interdefinitions of truth-functional operations, (1) and (2) imply that E is closed under truth-functional combinations of its own elements.

As in the urn example, probability revision will take the form of changes in the probabilities we assign to possible probabilistic structures of the world. These alternative structures are represented by elements of B . Since only the probabilities, not the nature, of these states are subject to change, the conditional probabilities of observables, given each element of B , will not change. In other words, the following property will hold for all probability functions p under consideration, all $b \in B$ and all $a_1, a_2 \in E$:

$$(p * a_1)(a_2 \mid b) = p(a_2 \mid b) \quad (2)$$

We can now introduce a model of probability revision that follows the pattern of reasoning introduced above in the discussion of the urn problem:¹¹

Definition 4 A *multistate model* of real-valued probability revision is a quadruple $\langle p, B, E, * \rangle$, in which p is a real-valued probability function, B a state catalogue for p , E a set of observables for p that is logically disjoint from B , and $*$ an operation of revision such that for all $a_1, a_2 \in E$ and $b \in B$:

If $p(a_1) = 0$, then:

$$(0) \quad p * a_1 = p$$

If $p(a_1) \neq 0$, then:

$$(1) \quad (p * a_1)(b) = p(b \mid a_1) = \frac{p(a_1 \& b)}{p(a_1)}$$

$$(2) \quad (p * a_1)(a_2 \mid b) = p(a_2 \mid b) = \frac{p(a_2 \& b)}{p(b)}$$

$$(3) \quad (p * a_1)(a_2) = \sum_{b \in B} (p * a_1)(b) \times (p * a_1)(a_2 \mid b) = \sum_{b \in B} \left(\frac{p(a_1 \& b)}{p(a_1)} \times \frac{p(a_2 \& b)}{p(b)} \right)$$

2.4 A Multistate Model for Hyperreal Probabilities

Let us now adapt the model developed in the previous subsection to hyperreal numbers. Definition 2 does not need any change to allow for hyperreal probabilities. However, the possibility of including states with infinitesimal probability in the state catalogue B is crucial for the purpose of this model. These are the states that the epistemic agent does not currently consider to be serious possibilities, but which she may possibly, on some future occasion, upgrade to serious possibilities. To illustrate this, suppose in our urn example that we repeatedly pull out a ball, note its colour, put it back, thoroughly shake the urn, again pull out a ball, etc. If we have done this one hundred times, and pulled out a yellow ball in each of these draws, then we are justified in assigning a real-valued non-zero probability to some previously disregarded

¹¹Two interesting comparisons can be drawn between this model and other accounts of epistemic change. First, the role of the state catalogue B can be compared to that of a belief base in dichotomous belief change. A belief base is a (usually finite) set of propositions whose logical closure is identical to the belief set. All changes are performed on the belief base, and the new belief set is obtained as the logical closure of the new belief base [11, 15]. This is analogous to the role of the state catalogue B in the present model, if we replace logical inference by the method for inferring probabilities described in Definition 4.

Secondly, the multistate model is closely related to models employing second-order probabilities [7, 17, 42]. In such models, there is a set $\mathfrak{P} = \{p_1, p_2, \dots\}$ of first order probability functions, all of which have the same domain, and a second-order probability function $\hat{p}$ that assigns non-zero probabilities to the elements of $\mathfrak{P}$. The probability of a proposition a is equal to its $\hat{p}$ -weighted first-order probability, i.e.:

$$p(a) = \sum_{p_k \in \mathfrak{P}} \hat{p}(p_k) \times p_k(a)$$

This approach can be translated directly into the present framework. For that purpose, we construct a state catalogue B with a one-to-one correspondence between elements $p_k \in \mathfrak{P}$ and states $b_k \in B$. Furthermore, we introduce a probability function p' such that (1) $p'(b_k) = \hat{p}(p_k)$ for all $b_k \in B$, and (2) $p'(a \mid b_k) = p_k(a)$ for all $b_k \in B$ and all propositions a in the domain of the elements of $\mathfrak{P}$.

states, such as a state in which all balls in the urn are yellow. For this to be possible, the state catalogue must contain such a state, although it was previously not considered to be a serious possibility. In our model, such a belief state initially had infinitesimal probability, but it can be raised to real-valued probability, depending on the information received.¹²

Just like Definition 2, Definition 3 can be used without changes in a model with hyperreal probabilities.

Definition 4, however, requires a rather thorough adaptation. We are aiming at a model in which elements of B that are currently not considered to be serious possibilities can be retained (with infinitesimal probabilities) for possible future invigoration. To achieve that, we need to modify the inputs of revision, and adjust the operation of revision accordingly. For a simple example, let $b \in B$ and $p(a_1 | b) = 0$. Then it follows from Definition 4 that $(p * a_1)(b) = 0$ and consequently $(p * a_1 * a_2 * \dots * a_n)(b) = 0$ for all series $a_2, \dots, a_n$ of propositions, i.e. b is lost for ever, which is exactly what we wanted to avoid.

To solve this, we need to replace the operation of revision employed in Definition 4, which assigns probability 1 to the new full belief a_1 , with an operation that assigns to it a probability $1 - \delta$, where δ is either 0 or an infinitesimal number. This requires that we change the format of inputs. Instead of performing revision by a proposition a , we will follow Jeffrey [24, pp. 164–183] in using inputs of the form $\langle s, a \rangle$, where a is a proposition and s its new probability after the operation has been performed. For our present purposes, we only need to perform revisions by inputs $\langle s, a \rangle$ such that $st(s) = 1$, i.e. we revise by $\langle 1 - \delta, a \rangle$, where δ is either 0 or a positive infinitesimal number. Instead of using standard conditionalization as a suboperation of revision, as in Definition 4, we will use Jeffrey conditionalization, i.e.:

$$p(e_2 | \langle s, e_1 \rangle) = s \times p(e_2 | e_1) + (1 - s) \times p(e_2 | \neg e_1) \quad (3)$$

This has the important consequence that when we revise by a sentence a , then we do not – as in standard Bayesian conditionalization – have to remove all traces of $\neg a$. To the contrary, if $\delta \neq 0$, then $\neg a$ is retained as an infinitesimal trace, along with infinitesimal traces of the conditional probabilities that have $\neg a$ as their antecedent. They can all be recovered to “reality” (i.e. real-valued probabilities) when that is needed.¹³

The following simplifying notation will be used, under the assumption that $st(\delta) = 0 \leq \delta$:

$$p *_\delta a = p * \langle 1 - \delta, a \rangle \quad (4)$$

$$p *_0 a = p * \langle 1, a \rangle \quad (5)$$

¹²Alternatively, we can model this process as the addition of a new element to the state catalogue. One way to do this was explored in considerable detail by Wenmackers and Romeijn [46]. A portion of the total probability is assigned to the “catch-all hypothesis” that none of the specific theories to which probability has been assigned is true. A new theory can then be introduced by receiving initial probability “shaven off” from the probability of the catch-all hypothesis. In their formal development of this approach, Wenmackers and Romeijn assigned real-valued probability to the catch-all hypothesis, but they mentioned that it could instead be assigned infinitesimal probability [46, 1248–1249].

¹³This use of infinitesimals is one of the methods mentioned by Skyrms [43] to avoid the irreversibility of Bayesian updating and to retain memory of $p(e | a)$ after revision by a .

The limiting clause (0) of Definition 4 will still apply, but in a slightly adjusted form:

$$\text{If } p(a_1) = 0, \text{ then } p *_{\delta} a_1 = p. \quad (6)$$

In other words, it is still impossible to revise by a sentence that has the antecedent probability 0. However, this limiting case will be less prominent in the hyperreal version of the model, since $p *_{\delta} a_1$ does not have to be equal to p in the important case when $\text{st}(p(a_1)) = 0 \neq p(a_1)$.

Clause (1) of Definition 4 has to be adjusted to Jeffrey conditionalization, thus:

$$\begin{aligned} (p *_{\delta} a_1)(b) &= (p * \langle 1 - \delta, a_1 \rangle)(b) \\ &= p(b \mid \langle 1 - \delta, a_1 \rangle) \\ &= (1 - \delta) \times p(b \mid a_1) + \delta \times p(b \mid \neg a_1) \\ &= (1 - \delta) \times \frac{p(a_1 \& b)}{p(a_1)} + \delta \times \frac{p(\neg a_1 \& b)}{p(\neg a_1)} \\ &= \frac{p(a_1 \& b)}{p(a_1)} + \delta \times \left(\frac{p(\neg a_1 \& b)}{p(\neg a_1)} - \frac{p(a_1 \& b)}{p(a_1)} \right) \end{aligned}$$

Revision in the multistate model takes the form of changes in the probabilities assigned to elements of B , which are (unchangeable) probabilistic states. Clause (2) ensures the permanence of these states, encoded in the constancy of their degrees of probabilistic support of different observables. This is an essential property of the multistate model, and clause (2) is not in need of any other modification than replacement of the revision operation $*$ by the more general $*_{\delta}$.

Clause (3) requires a change similar to that of clause (1):

$$\begin{aligned} (p *_{\delta} a_1)(a_2) &= \sum_{b \in B} (p * \langle 1 - \delta, a_1 \rangle)(b) \times (p * \langle 1 - \delta, a_1 \rangle)(a_2 \mid b) \\ &= \sum_{b \in B} \left(\frac{(1 - \delta) \times p(a_1 \& b)}{p(a_1)} + \frac{\delta \times p(\neg a_1 \& b)}{p(\neg a_1)} \right) \times \frac{p(a_2 \& b)}{p(b)} \\ &= \sum_{b \in B} \left(\frac{p(a_1 \& b)}{p(a_1)} + \delta \left(\frac{p(\neg a_1 \& b)}{p(\neg a_1)} - \frac{p(a_1 \& b)}{p(a_1)} \right) \right) \times \frac{p(a_2 \& b)}{p(b)} \end{aligned}$$

Finally, as the reader may have noticed, our new versions of clauses 1 and 3 are undefined if $p(a_1) = 1$, since they have $p(\neg a_1)$ in a denominator. Therefore, this case has to be excluded from these clauses, and dealt with by special postulation. In the real-valued model of Definition 4, it follows from clauses (1)-(3) that $p * a_1 = p$ if $p(a_1) = 1$.¹⁴ This is a sensible solution since a revision by a proposition that one is already unable to doubt should be vacuous and leave the belief state as it was. We can achieve this by extending clause (0) to this case as well:

$$\text{If } p(a_1) = 0 \text{ or } p(a_1) = 1, \text{ then } p *_{\delta} a_1 = p \quad (7)$$

Summarizing all this, we obtain the following multistate model of revision with hyperreal probabilities:

¹⁴To see this, note that in this case, $p(a_1 \& b) = p(b)$. It then follows from clause (1) that $(p * a_1)(b) = p(b)$ and from clause (3) that $(p * a_1)(a_2) = \sum_{b \in B} p(a_2 \& b) = p(a_2)$.

Definition 5 A *multistate model* of (hyperreal) probability revision is a quadruple $\langle p, B, E, * \rangle$, in which p is a hyperreal-valued probability function, B a state catalogue for p , E a set of observables for p that is logically disjoint from B , and $*$ an operation of revision such that for all $a_1, a_2 \in E$ and $b \in B$:

If $p(a_1) = 0$ or $p(a_1) = 1$, then:

$$(0) \quad p *_\delta a_1 = p$$

If $0 \neq p(a_1) \neq 1$, then:

$$(1) \quad (p *_\delta a_1)(b) = \frac{p(a_1 \& b)}{p(a_1)} + \delta \times \left(\frac{p(\neg a_1 \& b)}{p(\neg a_1)} - \frac{p(a_1 \& b)}{p(a_1)} \right)$$

$$(2) \quad (p *_\delta a_1)(a_2 | b) = p(a_2 | b) = \frac{p(a_2 \& b)}{p(b)}$$

$$(3) \quad (p *_\delta a_1)(a_2) = \sum_{b \in B} \left(\frac{p(a_1 \& b)}{p(a_1)} + \delta \left(\frac{p(\neg a_1 \& b)}{p(\neg a_1)} - \frac{p(a_1 \& b)}{p(a_1)} \right) \right) \times \frac{p(a_2 \& b)}{p(b)}$$

3 Properties of Probability Revision in the Multistate Model

The following observation confirms that if p satisfies the Kolmogorov axioms, then so does every new probability function that is obtainable by revision of p :¹⁵

Observation 1 Let p satisfy the Kolmogorov axioms. Then:

$$(1) \quad (p *_\delta a)(d) \geq 0$$

$$(2) \quad (p *_\delta a)(\top) = 1$$

(3) If $d_1, \dots, d_n$ are all mutually exclusive, then:

$$(p *_\delta a)(d_1 \vee \dots \vee d_n) = \sum_{i=1}^n (p *_\delta a)(d_i)$$

The following observation provides some convenient interrelations between the standard parts of revised hyperreal probabilities.

Observation 2 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision.

(1) Let $e_1, \dots, e_n$ be mutually exclusive elements of E , let $d \in E$, and let each of $\delta_1, \dots, \delta_n$ and δ be either 0 or a positive infinitesimal. Then:

$$p(e_1 \vee \dots \vee e_n) \times (p *_\delta (e_1 \vee \dots \vee e_n))(d) \approx \sum_{i=1}^n (p(e_i) \times (p *_\delta e_i)(d))$$

(2) Let a_1 and a_2 be elements of E , both with non-zero probabilities, and let each of δ_1 and δ_2 be either 0 or a positive infinitesimal. Then:

$$p(a_2) \times (p *_\delta a_2)(a_1) \approx p(a_1) \times (p *_\delta a_1)(a_2)$$

These are properties shared by standard Bayesian conditionalization (which is of course a special case of the multistate model, in which B has exactly one

¹⁵We assume a non-standard variant of the Kolmogorov axioms, in which probabilities are defined on the hyperreal rather than the real interval $[0, 1]$. As explained above in Section 1, this is possible since the domain of the probability function is finite.

element). The following conditions hold for Bayesian conditionalization, provided that $e_1, \dots, e_n$ are mutually exclusive:

$$p(e_1 \vee \dots \vee e_n) \times p(d \mid e_1 \vee \dots \vee e_n) = \sum_{i=1}^n (p(e_i) \times p(d \mid e_i)) \quad (8)$$

$$p(a_2) \times p(a_1 \mid a_2) = p(a_1 \& a_2) = p(a_1) \times p(a_2 \mid a_1) \quad (9)$$

The resources of this model allow us to divide sentences into an exhaustive set of five mutually exclusive epistemic classes:

- irreversible full belief ($p(a) = 1$)
- reversible full belief ($p(a) \lesssim 1$)
- irresolute belief ($0 \ll p(a) \ll 1$)
- reversible full disbelief ($0 \lesssim p(a)$)
- irreversible full disbelief ($0 = p(a)$)

Traditional (Bayesian) probability dynamics contains only the first, third, and fifth of these classes. Three of the classes have close analogues in dichotomous belief change, namely the second ($a \in K$), third ($a \notin K, \neg a \notin K$) and fourth ($\neg a \in K$).

The following two observations show what patterns of movement are possible among the five epistemic classes.

Observation 3 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision. Let $0 \leq \delta \approx 0$. Then:

- (1) If $p(a) = 1$ then $(p *_\delta d)(a) = 1$ for all propositions d .
- (2) If $p(a) = 0$ then $(p *_\delta d)(a) = 0$ for all propositions d .

Observation 4 There are multistate models of probability revision that satisfy each of the following conditions for some propositions a and d and some number δ with $\text{st}(\delta) = 0$:

- (1) $p(a) \lesssim 1$ and $(p *_\delta d)(a) = 1$
- (2) $p(a) \lesssim 1$ and $0 \ll (p *_\delta d)(a) \ll 1$
- (3) $p(a) \lesssim 1$ and $0 \lesssim (p *_\delta d)(a)$
- (4) $p(a) \lesssim 1$ and $0 = (p *_\delta d)(a)$
- (5) $0 \ll p(a) \ll 1$ and $(p *_\delta d)(a) = 1$
- (6) $0 \ll p(a) \ll 1$ and $(p *_\delta d)(a) \lesssim 1$
- (7) $0 \ll p(a) \ll 1$ and $0 \lesssim (p *_\delta d)(a)$
- (8) $0 \ll p(a) \ll 1$ and $0 = (p *_\delta d)(a)$
- (9) $0 \lesssim p(a)$ and $(p *_\delta d)(a) = 1$
- (10) $0 \lesssim p(a)$ and $(p *_\delta d)(a) \lesssim 1$
- (11) $0 \lesssim p(a)$ and $0 \ll (p *_\delta d)(a) \ll 1$
- (12) $0 \lesssim p(a)$ and $0 = (p *_\delta d)(a)$

Thus, once a proposition has been placed in either the first or the fifth class, no revision or series of revisions can remove it from there. On the other hand, from each of the other three classes, moves to all the other classes are possible. In Fig. 1, these relationships of movability are compared to the much less extensive set of moves that are possible in classical probabilistic frameworks with only real-valued probabilities.

Fig. 1 To the left, the three belief classes in the traditional approach to real-valued probabilities, and the two types of possible movements between these classes. To the right, the five belief classes in the multistage hyperreal model, and the twelve types of possible movements between them

Learning that an event has taken place can induce us to increase our previous estimate of its (prior) probability. This was illustrated in the urn example discussed in Section 2.2, where we noted an increase in the assessed probability of an event a from $p(a) = 0.5$ to $(p *_{\delta} a)(a) = 0.625$. Such changes appear to be common in practice, although they have sometimes been described as irrational [22]. They are difficult to represent in a standard Bayesian framework, since the most obvious representation $p(a \mid a)$ does not work for the purpose. (The value of that expression is 1 whenever a has non-zero probability.) In the multistate model they are represented by expressions such as $(p *_{\delta} a)(a)$, and typically, $(p *_{\delta} a)(a) \neq 1$. The following observation reports some results on this type of epistemic change.

Observation 5 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision, and let $\delta \approx 0$. Then:

- (1) $\text{st}(p(a)) \leq \text{st}((p *_{\delta} a)(a))$
- (2) If there is some d such that $(p *_{\delta} d)(a) \neq p(a)$, then $p(a) \ll (p *_{\delta} a)(a)$.

Observation 6 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision. Then:

- (1) It does not hold in general that if $d \vdash a$, then $(p *_{\delta} d)(a) \geq p(a)$
- (2) It does not hold in general that $(p *_{\delta} a)(a) \geq (p *_{\delta} d)(a)$.

Observation 7 $p(a) \times (p * a)(\neg a) \approx p(\neg a) \times (p * \neg a)(a)$

As noted above, the $*$ notation provides a convenient way to express iterated revision. The following theorem shows that iterated revision has interesting properties in the multistate model.

Theorem 1 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision. Let $b' \in B$ and $a_1, \dots, a_n, d \in E$, and let each of $\delta_1, \dots, \delta_n$ be either 0 or a positive infinitesimal number. Then:

- (1) If $0 \ll p(a_1 \& \dots \& a_n)$, then:

$$(p *_{\delta_1} a_1 \dots *_{\delta_n} a_n)(b') \approx \frac{p(b') \times p(a_1 \mid b') \times \dots \times p(a_n \mid b')}{\sum_{b \in B} p(b) \times p(a_1 \mid b) \times \dots \times p(a_n \mid b)}$$

- (2) If $0 \ll p(a_1 \& a_2)$, then:
 $(p *_{\delta_1} a_1 *_{\delta_2} a_2)(d) \approx (p *_{\delta_2} a_2 *_{\delta_1} a_1)(d)$

Hence, in the multistate model, probability revision by a series of compatible propositions is permutative.

It is essential for probabilistic learning that we can make several observations of what we consider to be the same phenomenon, such as several tosses of the same dice or several accidents of the same type. Such “duplicates” can be represented as follows:

Definition 6 Let $D = \{d_1, \dots, d_n\}$ be a finite set of distinct propositions. A *full specification* of D is a conjunction $e_1 \& \dots \& e_n$, in which each e_i is equal to either d_i or $\neg d_i$.

Definition 7 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision. A set $D \subseteq E$ is a *set of probabilistic duplicates* if and only if it holds for all $b \in B$ and all $a, a' \in D$ that if d is a full specification of $D \setminus \{a\}$ and d' a full specification of $D \setminus \{a'\}$, then

$$p(a \mid b \& d) = p(a' \mid b \& d')$$

We can use $(p[*_{\delta} a]^n)(b)$ as an abbreviated notation for $(p *_{\delta} a^1 *_{\delta} \dots *_{\delta} a^n)(b)$, where $b \in B$ and $\{a^1, \dots, a^n\}$ is a set of probabilistic duplicates of a .

Observation 8 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision. Let $a \in E$ be an observable that has probabilistic duplicates, and let b' be an element of B such that $p(a \mid b') > p(a \mid b)$ for all $b \in B \setminus \{b'\}$. Then $\text{st}((p[*_{\delta} a]^n)(b'))$ converges to 1 with increasing n .

For events whose combination has at most infinitesimal probability, the permutative result of Theorem 1 does not hold in general:

Observation 9 Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision.

- (1) Let $p(a_1 \& a_2) = 0$. It does not hold in general that $(p *_{\delta_1} a_1 *_{\delta_2} a_2)(d) \approx (p *_{\delta_2} a_2 *_{\delta_1} a_1)(d)$.
 (2) Let $0 \lesssim p(a_1 \& a_2)$. It does not hold in general that $(p *_{\delta_1} a_1 *_{\delta_2} a_2)(d) \approx (p *_{\delta_2} a_2 *_{\delta_1} a_1)(d)$.

In the example used in the proof of Observation 9, the iterated operation gives preference to the most recent information.¹⁶ The common models of dichotomous belief change have the same tendency, but to an even higher degree; they always give priority to the most recent input and delete whatever needs to be removed in order to

¹⁶ $p *_{\delta_1} a_1 *_{\delta_2} a_2$ favours b_2 in which $\neg a_1 \& a_2$ always holds over b_1 in which $a_1 \& \neg a_2$ always holds. In contrast, $p *_{\delta_2} a_2 *_{\delta_1} a_1$ has the opposite priority.

accommodate it. [16, 21] This is a feature of the multistate model that seems to be worth a more thorough investigation.

4 A Derived Logic of Full Beliefs

One obvious way to connect probabilistic and dichotomous belief change models is to treat the “top” of a standard probability function, i.e. the set of sentences to which it assigns probability 1, as a (dichotomous) belief set. However, this approach has the serious drawback that in the standard model, this top will act as a black hole, a belief that has reached it can never be removed from it again. A much more reasonable model can be obtained by instead using the “soft top” of the probabilistic belief states in the (hyperreal) multistate model. The soft top consists of the propositions whose probability has the standard part 1. It corresponds to the two uppermost levels in the right-hand diagram in Fig. 1, but for most purposes only the lowest of these two levels should be used for empirical propositions. Since different infinitesimal numbers, as well as 0, can serve as the index δ in the revision operation $*_{\delta}$, different revisions by one and the same proposition are available.¹⁷

In this section, the focus will be on the derived system of dichotomous belief change that can be obtained from the multistate model in this way. To begin with, we need to confirm that the soft top is really a (logically closed) belief set.

Definition 8 Let p be the probability function of a multistate model $\langle p, B, E, * \rangle$ of probability revision. The set

$$\llbracket p \rrbracket = \{a \mid \text{st}(p(a)) = 1\}$$

is the *belief set* generated by the probability function p .

Observation 10 Let p be the probability function of a multistate model $\langle p, B, E, * \rangle$ of probability revision. Then:

$$\llbracket p \rrbracket = \text{Cn}(\llbracket p \rrbracket) \text{ (top closure)}$$

The most influential approach to dichotomous belief change is the AGM model, which is characterized by a set of postulates. Although these postulates are far from uncontroversial [18, pp. 27-42] and [5, pp. 41-47], it is common practice to compare other dichotomous belief change models to them. The postulates come in two groups:¹⁸

The six basic AGM postulates:

$$K * a = \text{Cn}(K * a) \text{ (closure)}$$

$$K * a \subseteq K + a \text{ (inclusion)}$$

$$\text{If } \neg a \notin K, \text{ then } K + a \subseteq K * a. \text{ (vacuity)}$$

$$a \in K * a \text{ (success)}$$

¹⁷If the framework contains revisions by inputs $\langle s, a \rangle$ with $s \lesssim 1$, then these operations can be seen from the dichotomous viewpoint as operations that strengthen or weaken the credibility of the proposition a without changing its (dichotomous) belief status. Such operations have been discussed in the belief revision literature, see [4, 28, 29].

¹⁸ $K + a$ is an abbreviation of $\text{Cn}(K \cup \{a\})$.

If $a \leftrightarrow a'$ is a logical truth, then $K * a = K * a'$. (extensionality)

If a is consistent, then so is $K * a$. (consistency)

The two supplementary AGM postulates

$K * (a \& a') \subseteq (K * a) + a'$ (superexpansion)

If $\neg a' \notin K * a$, then $(K * a) + a' \subseteq K * (a \& a')$. (subexpansion)

The following observation shows that four of the six basic AGM postulates hold for revision in the multistate model:

Observation 11 *Let $\mathbf{p}$ be the probability function of a multistate model $\langle \mathbf{p}, B, E, * \rangle$ of probability revision. Then:*

- (1) $\llbracket \mathbf{p} *_\delta a \rrbracket = \text{Cn}(\llbracket \mathbf{p} *_\delta a \rrbracket)$ (closure)
- (2) $\llbracket \mathbf{p} *_\delta a \rrbracket \subseteq \text{Cn}(\llbracket \mathbf{p} \rrbracket \cup \{a\})$ (inclusion)
- (3) *It does not hold in general that if $\neg a \notin \llbracket \mathbf{p} \rrbracket$, then there is some δ such that $\text{Cn}(\llbracket \mathbf{p} \rrbracket \cup \{a\}) \subseteq \llbracket \mathbf{p} *_\delta a \rrbracket$. (“vacuity”)*
- (4) *It does not hold in general that there is some δ with $0 \leq \delta \approx 0$ and $a \in \llbracket \mathbf{p} *_\delta a \rrbracket$. (“success”)*
- (5) *If $\vdash a_1 \leftrightarrow a_2$, then $\llbracket \mathbf{p} *_\delta a_1 \rrbracket = \llbracket \mathbf{p} *_\delta a_2 \rrbracket$ (extensionally)*
- (6) $\llbracket \mathbf{p} *_\delta a \rrbracket \not\vdash \perp$ (strong consistency)¹⁹

The success postulate, which fails for this operation, is probably the most criticized among the AGM postulates for revision. Quite a few authors have found it utterly unrealistic that an operation of revision should accept any new information that is received. There is also a considerable literature on non-prioritized revision, by which is meant revision which does not always accept its inputs. (See Fermé and Hansson [5, pp. 65–68] for an overview.)

Several weaker versions of the success postulate have been proposed. The following two are satisfied by several operations of non-prioritized revision:

Either $K * p \vdash p$ or $K * p = K$. (relative success)

If $K * q \vdash p$, then $K * p \vdash p$. (regularity)

However, neither of these holds for the multistate revision operation:

Observation 12 *Let $\mathbf{p}$ be the probability function of a multistate model $\langle \mathbf{p}, B, E, * \rangle$ of probability revision, and let $a \in E$. Then:*

- (1) *It does not hold in general that there is some δ such that $a \in \llbracket \mathbf{p} *_\delta a \rrbracket$ or $\llbracket \mathbf{p} *_\delta a \rrbracket = \llbracket \mathbf{p} \rrbracket$.*
- (2) *It does not hold in general that there is some δ such that: If $a \in \llbracket \mathbf{p} *_\delta d \rrbracket$, then $a \in \llbracket \mathbf{p} *_\delta a \rrbracket$.*

Strong consistency, which was shown in Observation 11 to hold in the multistate model, is significantly stronger than the common consistency postulate of the AGM framework (which would in our notation be expressed as: If $a \not\vdash \perp$, then $\llbracket \mathbf{p} *_\delta a \rrbracket \not\vdash \perp$).

¹⁹Hansson [12, p. 344] and [14, p. 154].

The strong consistency postulate is incompatible with the success postulate, since the former forbids but the latter requires that $\llbracket p *_\delta \perp \rrbracket \vdash \perp$.

In the presence of closure, the vacuity postulate,

$$\text{If } \neg a \notin \llbracket p \rrbracket, \text{ then } \text{Cn}(\llbracket p \rrbracket \cup \{a\}) \subseteq \llbracket p *_\delta a \rrbracket,$$

is equivalent with the conjunction of the following two conditions:

- (i) If $\neg a \notin \llbracket p \rrbracket$, then $a \in \llbracket p *_\delta a \rrbracket$
- (ii) If $\neg a \notin \llbracket p \rrbracket$, then $\llbracket p \rrbracket \subseteq \llbracket p *_\delta a \rrbracket$ (preservation)²⁰

On closer inspection it turns out that the reason vacuity does not hold for our probability-based operation of revision is that (i) does not hold. In contrast, (ii) holds:

Observation 13 *Let p be the probability function of a multistate model $\langle p, B, E, * \rangle$ of probability revision. Then:*

$$\text{If } \neg a \notin \llbracket p \rrbracket, \text{ then } \llbracket p \rrbracket \subseteq \llbracket p *_\delta a \rrbracket \text{ (preservation)}$$

Since (i) follows from success, it would make no difference to replace vacuity by preservation (ii) in the list of basic AGM axioms. After such a modification, success would be the only of these postulates not satisfied in the multistate model.

Neither of the two supplementary postulates of the AGM model holds for the multistate operation.

Observation 14 *Let $\langle p, B, E, * \rangle$ be a multistate model of probability revision.*

- (1) *It does not hold in general that there are δ and δ' such that $\llbracket p *_\delta (a_1 \& a_2) \rrbracket \subseteq \text{Cn}(\llbracket p *_\delta a_1 \rrbracket \cup \{a_2\})$. (“superexpansion”)*
- (2) *It does not hold in general that if $\neg a_2 \notin \llbracket p *_\delta a_1 \rrbracket$, then there is some δ' such that $\text{Cn}(\llbracket p *_\delta a_1 \rrbracket \cup \{a_2\}) \subseteq \llbracket p *_\delta (a_1 \& a_2) \rrbracket$. (“subexpansion”)*

As emphasized in Section 1, a major purpose of the present framework is to allow for reversible movements between the set of irresolute beliefs and the set of full beliefs. We have investigated operations that transfer a proposition from the set of resolute beliefs to that of full beliefs. For changes in the opposite direction we will need other types of operations, namely operations designed to decrease the probability of a given proposition. In the language of dichotomous belief change, such operations are called contractions. The construction of suitable such operations and the investigation of their relations to other operations of contraction are left for a later occasion.

Acknowledgements Open access funding provided by Royal Institute of Technology. I would like to thank Branden Fitelson, Wlodek Rabinowicz, and the other participants at the workshop *Cognitive values: Theoretical and Practical Problems* at the Royal Swedish Academy of Letters, History and Antiquities in Stockholm on June 10–11, 2019, for most valuable comments. I am also grateful to an anonymous referee for unusually helpful proposals for improvements of the text.

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

²⁰Gärdenfors [8, p. 84].

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/>.

Appendix A: A Brief Introduction to Hyperreal Numbers

Readers wanting an introduction to the arithmetic of hyperreals and infinitesimals are referred to Keisler [25].

The number system of hyperreal numbers (R^*) is an extension of the set R of real numbers. It consists of both finite and infinite numbers, but our focus will be on the finite numbers. The hyperreal finite numbers consist of (1) the real numbers and (2) other numbers on an extended number line, which are posited between the real numbers. The positive hyperreal numbers that are infinitely close to 0 (larger than 0 but smaller than all positive real numbers) are the positive infinitesimals. The negative infinitesimals are the numbers that are smaller than 0 but larger than all negative real numbers.

Two hyperreal numbers s and t are infinitesimally close (denoted $s \approx t$) if and only if their difference $s - t$ is infinitesimal. The relation $\approx$ is an equivalence relation. Each finite hyperreal number is infinitely close to exactly one real number, which is called its standard part. The standard part of the number t is denoted $\text{st}(t)$. Standard parts satisfy the following rules.

$$\text{st}(-s) = -\text{st}(s)$$

$$\text{st}(s + t) = \text{st}(s) + \text{st}(t)$$

$$\text{st}(s - t) = \text{st}(s) - \text{st}(t)$$

$$\text{st}(s \times t) = \text{st}(s) \times \text{st}(t)$$

$$\text{If } \text{st}(t) \neq 0 \text{ then } \text{st}(s/t) = \text{st}(s)/\text{st}(t)$$

The set of hyperreal numbers satisfies the same algebraic laws as the real numbers. Let δ and ϵ be infinitesimals, and let s and t be finite numbers that are not infinitesimals. Then:

The following are infinitesimals: $\delta + \epsilon$, $\delta\epsilon$, $s\delta$ and δ/s .

The following are finite and not infinitesimal: $s + \epsilon$, st and s/t .

The following is finite and may or may not be infinitesimal: $s + t$

The following may be infinite or finite, and in the latter case it may or may not be infinitesimal: δ/ϵ .

Appendix B: Proofs

Proof of Observation 1 If $\text{p}(a) = 0$ or $\text{p}(a) = 1$, then all three properties follow directly from clause (0) of Definition 5. If $0 \neq \text{p}(a) \neq 1$, then:

Part 1: In clause (1) of the definition, rewrite the right-hand side as follows:

$$(1 - \delta) \times \frac{\text{p}(a_1 \& b)}{\text{p}(a_1)} + \delta \times \frac{\text{p}(\neg a_1 \& b)}{\text{p}(\neg a_1)}$$

Since δ , $1 - \delta$, $p(a_1 \& b)/p(a_1)$, and $p(\neg a_1 \& b)/p(\neg a_1)$ are all non-negative, so is the whole expression. The non-negativity of the right-hand side of clause (3) follows in the same way.

Part 2: Clause (3) of Definition 5 yields:

$$\begin{aligned} (p *_{\delta} a)(\top) &= (1 - \delta) \times \left(\sum_{b \in B} \frac{p(a \& b) \times p(\top \& b)}{p(a) \times p(b)} \right) + \delta \times \left(\sum_{b \in B} \frac{p(\neg a \& b) \times p(\top \& b)}{p(\neg a) \times p(b)} \right) \\ &= \left(\frac{1 - \delta}{p(a)} \times \sum_{b \in B} p(a \& b) \right) + \left(\frac{\delta}{p(\neg a)} \times \sum_{b \in B} p(\neg a \& b) \right) \\ &= \left(\frac{(1 - \delta)}{p(a)} \times p(a) \right) + \left(\frac{\delta}{p(\neg a)} \times p(\neg a) \right) = 1 \end{aligned}$$

Part 3: Using clause (3) of Definition 5, we obtain:

$$\begin{aligned} (p *_{\delta} a)(d_1 \vee \dots \vee d_n) &= (1 - \delta) \times \left(\sum_{b \in B} \frac{p(a \& b) \times p((d_1 \vee \dots \vee d_n) \& b)}{p(a) \times p(b)} \right) \\ &\quad + \delta \times \left(\sum_{b \in B} \frac{p(\neg a \& b) \times p((d_1 \vee \dots \vee d_n) \& b)}{p(\neg a) \times p(b)} \right) \\ &= (1 - \delta) \times \left(\sum_{b \in B} \sum_{i=1}^n \frac{p(a \& b) \times p(d_i \& b)}{p(a) \times p(b)} \right) \\ &\quad + \delta \times \left(\sum_{b \in B} \sum_{i=1}^n \frac{p(\neg a \& b) \times p(d_i \& b)}{p(\neg a) \times p(b)} \right) \\ &= \sum_{i=1}^n \left((1 - \delta) \times \left(\sum_{b \in B} \frac{p(a \& b) \times p(d_i \& b)}{p(a) \times p(b)} \right) \right. \\ &\quad \left. + \delta \times \left(\sum_{b \in B} \frac{p(\neg a \& b) \times p(d_i \& b)}{p(\neg a) \times p(b)} \right) \right) \\ &= \sum_{i=1}^n (p * \langle 1 - \delta, a \rangle)(d_i) = \sum_{i=1}^n (p *_{\delta} a)(d_i) \end{aligned}$$

□

Proof of Observation 2 Part 1: This follows directly if $p(e_1 \vee \dots \vee e_n) = 0$. If $p(e_1 \vee \dots \vee e_n) \neq 0$, then apply clause (3) of Definition 5. We obtain:

$$\begin{aligned} &p(e_1 \vee \dots \vee e_n) \times (p *_{\delta} (e_1 \vee \dots \vee e_n))(d) \\ &\approx p(e_1 \vee \dots \vee e_n) \times \sum_{b \in B} \left(\frac{p((e_1 \vee \dots \vee e_n) \& b)}{p(e_1 \vee \dots \vee e_n)} \times \frac{p(d \& b)}{p(b)} \right) \\ &= \sum_{b \in B} \left(p((e_1 \vee \dots \vee e_n) \& b) \times \frac{p(d \& b)}{p(b)} \right) \end{aligned}$$

$$\begin{aligned}
&= \sum_{b \in B} \sum_{i=1}^n \left(p(e_i \& b) \times \frac{p(d \& b)}{p(b)} \right) \\
&= \sum_{i=1}^n \sum_{b \in B} \left(p(e_i \& b) \times \frac{p(d \& b)}{p(b)} \right) \\
&= \sum_{i=1}^n \left(p(e_i) \times \sum_{b \in B} \left(\frac{p(e_i \& b)}{p(e_i)} \times \frac{p(d \& b)}{p(b)} \right) \right) \\
&\approx \sum_{i=1}^n (p(e_i) \times (p *_{\delta_i} e_i)(d))
\end{aligned}$$

Part 2:

$$\begin{aligned}
&p(a_2) \times (p *_{\delta_2} a_2)(a_1) \\
&\approx p(a_2) \times \left(\sum_{b \in B} \frac{p(a_2 \& b)}{p(a_2)} \times \frac{p(a_1 \& b)}{p(b)} \right) \quad (\text{Definition 5, clause (3)}) \\
&= p(a_1) \times \left(\sum_{b \in B} \frac{p(a_1 \& b)}{p(a_1)} \times \frac{p(a_2 \& b)}{p(b)} \right) \\
&\approx p(a_1) \times (p *_{\delta_1} a_1)(a_2)
\end{aligned}$$

□

Proof of Observation 3 Part 1: If $p(d) = 0$ or $p(d) = 1$, then the desired result follows directly from clause (0) of Definition 5. If $0 \neq p(d) \neq 1$, then we use clause (3) and obtain:

$$\begin{aligned}
(p *_{\delta} d)(a) &= \left(\frac{1 - \delta}{p(d)} \times \sum_{b \in B} \frac{p(d \& b) \times p(a \& b)}{p(b)} \right) \\
&+ \left(\frac{\delta}{p(\neg d)} \times \sum_{b \in B} \frac{p(\neg d \& b) \times p(a \& b)}{p(b)} \right) \\
&= \left(\frac{1 - \delta}{p(d)} \times \sum_{b \in B} p(d \& b) \right) + \left(\frac{\delta}{p(\neg d)} \times \sum_{b \in B} p(\neg d \& b) \right) \\
&= \left(\frac{1 - \delta}{p(d)} \times p(d) \right) + \left(\frac{\delta}{p(\neg d)} \times p(\neg d) \right) = 1
\end{aligned}$$

Part 2: If $p(d) = 0$ or $p(d) = 1$, then the desired result follows directly from clause (0) of Definition 5. If $0 \neq p(d) \neq 1$, then we use clause (3) and obtain:

$$\begin{aligned} (p *_{\delta} d)(a) &= \left(\frac{1 - \delta}{p(d)} \times \sum_{b \in B} \frac{p(d \& b) \times p(a \& b)}{p(b)} \right) \\ &\quad + \left(\frac{\delta}{p(\neg d)} \times \sum_{b \in B} \frac{p(\neg d \& b) \times p(a \& b)}{p(b)} \right) \\ &= 0 \quad (p(a \& b) = 0 \text{ for all } b \in B, \text{ due to } p(a) = 0) \end{aligned}$$

□

Proof of Observation 4

Parts 1, 5, and 9: Let $B = \{b_1, b_2\}$, $p(a \mid b_1) = 1$, $p(a \mid b_2) = 0$, and $p(b_1) = p(a)$. Setting $d = a$ and $\delta = 0$, we obtain from clause (3) of Definition 5:

$$\begin{aligned} (p *_{\delta} a)(a) &= \left(\frac{p(a \& b_1)}{p(a)} \times \frac{p(a \& b_1)}{p(b_1)} \right) + \left(\frac{p(a \& b_2)}{p(a)} \times \frac{p(a \& b_2)}{p(b_2)} \right) \\ &= 1 \quad (\text{Since } p(a \& b_1) = p(a) = p(b_1) \text{ and } p(a \& b_2) = 0.) \end{aligned}$$

Parts 2 and 11: Let $B = \{b_1, b_2\}$, $p(b_1) = p(a)$, $p(b_2) = 1 - p(a)$, $p(a \mid b_1) = 1$, $p(a \mid b_2) = 0$, and $p(d \& b_1) = p(d \& b_2) \neq 0$. It follows from clause (3) of Definition 5 that:

$$\begin{aligned} (p *_{\delta} d)(a) &= \left(\frac{p(d \& b_1)}{p(d)} \times \frac{p(a \& b_1)}{p(b_1)} \right) + \left(\frac{p(d \& b_2)}{p(d)} \times \frac{p(a \& b_2)}{p(b_2)} \right) \\ &= \frac{p(d \& b_1)}{p(d)} \quad (\text{Since } p(a \& b_1) = p(b_1) \text{ and } p(a \& b_2) = 0) \\ &= \frac{1}{2} \end{aligned}$$

Parts 3 and 7: Let $B = \{b_1, b_2\}$, $p(b_1) = p(a)$, $p(b_2) = 1 - p(a)$, $p(a \mid b_1) = 1$, and $p(a \mid b_2) = 0$. Let $0 \lesssim \delta$ and $d = \neg a$. It follows from clause (3) of Definition 5 that:

$$\begin{aligned} (p *_{\delta} \neg a)(a) &= \left(\frac{p(\neg a \& b_1)}{p(\neg a)} + \delta \left(\frac{p(a \& b_1)}{p(a)} - \frac{p(\neg a \& b_1)}{p(\neg a)} \right) \right) \times \frac{p(a \& b_1)}{p(b_1)} \\ &\quad + \left(\frac{p(\neg a \& b_2)}{p(\neg a)} + \delta \left(\frac{p(a \& b_2)}{p(a)} - \frac{p(\neg a \& b_2)}{p(\neg a)} \right) \right) \times \frac{p(a \& b_2)}{p(b_2)} \\ &= \delta \times \frac{p(a \& b_1) \times p(a \& b_1)}{p(a) \times p(b_1)} \quad (\text{Since } p(\neg a \& b_1) = p(a \& b_2) = 0) \\ &= \delta \quad (\text{Since } p(a \& b_1) = p(b_1) = p(a)) \end{aligned}$$

Parts 4, 8, and 12: Let $B = \{b_1, b_2\}$, $p(b_1) = p(a)$, $p(b_2) = 1 - p(a)$, $p(a \mid b_1) = 1$, $p(a \mid b_2) = 0$. Let $d = \neg a$ and $\delta = 0$. It follows from clause (3) of Definition 5 that:

$$(p *_{\neg a})(a) = \left(\frac{p(\neg a \& b_1)}{p(\neg a)} \times \frac{p(a \& b_1)}{p(b_1)} \right) + \left(\frac{p(\neg a \& b_2)}{p(\neg a)} \times \frac{p(a \& b_2)}{p(b_2)} \right) = 0$$

(Since $p(\neg a \& b_1) = 0$ and $p(a \& b_2) = 0$.)

Parts 6 and 10: Let $B = \{b_1, b_2\}$, $p(b_1) = p(a)$, $p(b_2) = 1 - p(a)$, $p(a \mid b_1) = 1$, and $p(a \mid b_2) = 0$. Let $d = a$ and let $0 \lesssim \delta$. Then:

$$\begin{aligned} (p *_{\delta} d)(a) &= (p * (1 - \delta, a))(a) \\ &= \left(\frac{p(a \& b_1)}{p(a)} + \delta \left(\frac{p(\neg a \& b_1)}{p(\neg a)} - \frac{p(a \& b_1)}{p(a)} \right) \right) \times \frac{p(a \& b_1)}{p(b_1)} \\ &\quad + \left(\frac{p(a \& b_2)}{p(a)} + \delta \left(\frac{p(\neg a \& b_2)}{p(\neg a)} - \frac{p(a \& b_2)}{p(a)} \right) \right) \times \frac{p(a \& b_2)}{p(b_2)} \\ &\quad \text{(Clause 3 of Definition 5)} \\ &= 1 - \delta \quad \text{(Since } p(a \& b_1) = p(a) = p(b_1) \text{ and } p(\neg a \& b_1) = p(a \& b_2) = 0.) \end{aligned}$$

□

Lemma 1 *Let $x_1, \dots, x_n$ and $y_1, \dots, y_n$ be positive numbers. Then:*

$$\frac{x_1^2}{y_1} + \dots + \frac{x_n^2}{y_n} \geq \frac{(x_1 + \dots + x_n)^2}{y_1 + \dots + y_n}$$

Proof of Lemma 1 We first prove the case when $n = 2$:

$$\begin{aligned} (x_1 y_2 - x_2 y_1)^2 &\geq 0 \\ x_1^2 y_2^2 + x_2^2 y_1^2 &\geq 2x_1 x_2 y_1 y_2 \\ x_1^2 y_1 y_2 + x_1^2 y_2^2 + x_2^2 y_1^2 + x_2^2 y_1 y_2 &\geq x_1^2 y_1 y_2 + x_2^2 y_1 y_2 + 2x_1 x_2 y_1 y_2 \\ x_1^2 y_2 (y_1 + y_2) + x_2^2 y_1 (y_1 + y_2) &\geq y_1 y_2 (x_1^2 + x_2^2 + 2x_1 x_2) \\ \frac{x_1^2}{y_1} + \frac{x_2^2}{y_2} &\geq \frac{(x_1 + x_2)^2}{y_1 + y_2} \end{aligned}$$

Next, we assume that this holds for n . We then have:

$$\frac{x_1^2}{y_1} + \dots + \frac{x_n^2}{y_n} + \frac{(x_{n+1})^2}{y_{n+1}} \geq \frac{(x_1 + \dots + x_n)^2}{y_1 + \dots + y_n} + \frac{(x_{n+1})^2}{y_{n+1}} \geq \frac{(x_1 + \dots + x_n + x_{n+1})^2}{y_1 + \dots + y_n + y_{n+1}}$$

□

Proof of Observation 5 Part 1: If $p(a) = 0$ or $p(a) = 1$, then $p * a = p$ and consequently $(p * a)(a) = p(a)$. If $0 \neq p(a) \neq 1$, then Lemma 1 yields:

$$\begin{aligned} \frac{p(a \& b_1)^2}{p(b_1)} + \dots + \frac{p(a \& b_n)^2}{p(b_n)} &\geq \frac{(p(a \& b_1) + \dots + p(a \& b_n))^2}{p(b_1) + \dots + p(b_n)} \\ \frac{p(a \& b_1)^2}{p(b_1)} + \dots + \frac{p(a \& b_n)^2}{p(b_n)} &\geq \frac{p(a)^2}{1} \\ \frac{p(a \& b_1)^2}{p(a)p(b_1)} + \dots + \frac{p(a \& b_n)^2}{p(a)p(b_n)} &\geq p(a) \\ \text{st}((p *_{\delta} a)(a)) &\geq \text{st}(p(a)) \end{aligned}$$

Part 2: Let $(p *_{\delta} d)(a) \not\approx p(a)$. First we note that if $p(a | b) \approx p(a)$ for all $b \in B$, then $(p *_{\delta} d)(a) \approx p(a)$. Thus there are elements b of B such that $p(a | b) \not\approx p(a)$. There are three options:

Case (i): For all b with $p(a) \ll p(a | b)$ we have:

$$\begin{aligned} p(a) &\ll \frac{p(a \& b)}{p(b)} \\ p(b) &\ll \frac{p(a \& b)}{p(a)} \\ p(b) &\ll p(b | a) \\ p(b) &\ll (p *_{\delta} a)(b) \end{aligned} \quad \text{(Clause (1) of Definition 5)}$$

Case (ii): For all b with $p(a) \approx p(a | b)$ we similarly obtain $p(b) \approx (p *_{\delta} a)(b)$.

Case (iii): For all b with $p(a | b) \ll p(a)$ we similarly obtain $(p *_{\delta} a)(b) \ll p(b)$.

It now follows from clause (3) of Definition 5 that $p(a) \ll (p *_{\delta} a)(a)$. $\square$

Proof of Observation 6 Part 1: Let $B = \{b_1, b_2\}$, let $a, d \in E$ and $d \vdash a$, and let p satisfy the following conditions: $p(b_1) = 0.5$, $p(b_1 \& a) = 0.4$, $p(b_1 \& a \& d) = 0.1$, $p(b_2 \& a) = 0.3$, and $p(b_2 \& a \& d) = 0.2$. Then $p(a) = 0.7$, and:

$$\begin{aligned} (p *_{\delta} d)(a) &= \frac{p(d \& b_1) \times p(a \& b_1)}{p(d) \times p(b_1)} + \frac{p(d \& b_2) \times p(a \& b_2)}{p(d) \times p(b_2)} \\ &= \frac{0.1 \times 0.4}{0.3 \times 0.5} + \frac{0.2 \times 0.3}{0.3 \times 0.5} = \frac{2}{3} < 0.7. \end{aligned}$$

Part 2: Let $B = \{b_1, b_2\}$, let $a, d \in E$, and let p satisfy the following conditions: $p(b_1) = 0.5$, $p(a \& b_1) = 0.375$, $p(d \& b_1) = 0.5$, $p(a \& b_2) = 0.125$, and $p(d \& b_2) = 0$. Then:

$$(p *_{\delta} a)(a) = \frac{p(a \& b_1)^2}{p(a) \times p(b_1)} + \frac{p(a \& b_2)^2}{p(a) \times p(b_2)} = \frac{0.375^2}{0.5^2} + \frac{0.125^2}{0.5^2} = 0.625$$

and:

$$\begin{aligned} (p *_{\delta} d)(a) &= \frac{p(d \& b_1) \times p(a \& b_1)}{p(d) \times p(b_1)} + \frac{p(d \& b_2) \times p(a \& b_2)}{p(d) \times p(b_2)} \\ &= \frac{0.5 \times 0.375}{0.5^2} + 0 = 0.75 \end{aligned} \quad \square$$

Proof of Observation 7 If $p(a) = 1$, then $p *_{\delta} a = p$, and $(p *_{\delta} a)(\neg a) = p(\neg a) = 0$, and the desired result follows directly. The case when $p(a) = 0$ can be proved analogously. If $0 \neq p(a) \neq 1$, then:

$$\begin{aligned}
 & p(a) \times (p * a)(\neg a) \approx \\
 & \approx p(a) \times \sum_{b \in B} \frac{p(a \& b) \times p(\neg a \& b)}{p(a) \times p(b)} \quad (\text{Definition 5, clause 3}) \\
 & = p(\neg a) \times \sum_{b \in B} \frac{p(\neg a \& b) \times p(a \& b)}{p(\neg a) \times p(b)} \\
 & \approx p(\neg a) \times (p * \neg a)(a)
 \end{aligned}$$

□

Proof of Theorem 1 Part 1: By induction. Initial step:

$$\begin{aligned}
 (p *_{\delta_1} a_1)(b') & \approx \frac{p(a_1 \& b')}{p(a_1)} \\
 & = \frac{p(b') \times p(a_1 | b')}{\sum_{b \in B} p(b) \times p(a_1 | b)}
 \end{aligned}$$

Induction step: We assume that

$$(p *_{\delta_1} a_1 *_{\delta_2} \dots *_{\delta_{n-1}} a_{n-1})(b') \approx \frac{p(b') \times p(a_1 | b') \times \dots \times p(a_{n-1} | b')}{\sum_{b \in B} p(b) \times p(a_1 | b) \times \dots \times p(a_{n-1} | b)}$$

We write $p_{n-1} = p *_{\delta_1} a_1 *_{\delta_2} \dots *_{\delta_{n-1}} a_{n-1}$ and $p_n = p *_{\delta_1} a_1 *_{\delta_2} \dots *_{\delta_n} a_n$. Then:

$$\begin{aligned}
 p_n(b') & = (p_{n-1} *_{\delta_n} a_n)(b') \\
 & \approx \frac{p_{n-1}(a_n \& b')}{p_{n-1}(a_n)} \\
 & = \frac{p_{n-1}(a_n \& b')}{\sum_{b \in B} p_{n-1}(a_n \& b)} \\
 & = \frac{p_{n-1}(b') \times p_{n-1}(a_n | b')}{\sum_{b \in B} p_{n-1}(b) \times p_{n-1}(a_n | b)} \\
 & = \frac{p_{n-1}(b') \times p(a_n | b')}{\sum_{b \in B} p_{n-1}(b) \times p(a_n | b)}
 \end{aligned}$$

With the induction hypothesis, this yields:

$$\begin{aligned}
 & \frac{\frac{\mathbf{p}(b') \times \mathbf{p}(a_1 | b') \times \dots \times \mathbf{p}(a_{n-1} | b')}{\sum_{b \in B} (\mathbf{p}(b) \times \mathbf{p}(a_1 | b) \times \dots \times \mathbf{p}(a_{n-1} | b))} \times \mathbf{p}(a_n | b')}{\sum_{b'' \in B} \left(\frac{\mathbf{p}(b'') \times \mathbf{p}(a_1 | b'') \times \dots \times \mathbf{p}(a_{n-1} | b'')}{\sum_{b \in B} (\mathbf{p}(b) \times \mathbf{p}(a_1 | b) \times \dots \times \mathbf{p}(a_{n-1} | b))} \times \mathbf{p}(a_n | b'') \right)} \\
 & \approx \frac{\frac{\mathbf{p}(b') \times \mathbf{p}(a_1 | b') \times \dots \times \mathbf{p}(a_n | b')}{\sum_{b \in B} (\mathbf{p}(b) \times \mathbf{p}(a_1 | b) \times \dots \times \mathbf{p}(a_{n-1} | b))}}{\frac{\sum_{b'' \in B} (\mathbf{p}(b'') \times \mathbf{p}(a_1 | b'') \times \dots \times \mathbf{p}(a_n | b''))}{\sum_{b \in B} (\mathbf{p}(b) \times \mathbf{p}(a_1 | b) \times \dots \times \mathbf{p}(a_{n-1} | b))}} \\
 & = \frac{\mathbf{p}(b') \times \mathbf{p}(a_1 | b') \times \dots \times \mathbf{p}(a_n | b')}{\sum_{b \in B} (\mathbf{p}(b) \times \mathbf{p}(a_1 | b) \times \dots \times \mathbf{p}(a_n | b))}
 \end{aligned}$$

Part 2:

$$(\mathbf{p} *_{\delta_1} a_1 *_{\delta_2} a_2)(d) \approx \sum_{b \in B} ((\mathbf{p} *_{\delta_1} a_1 *_{\delta_2} a_2)(b) \times \mathbf{p}(d | b)) \quad (\text{Definition 5, clauses (1) and (3)})$$

$$\begin{aligned}
 & \approx \sum_{b' \in B} \frac{\mathbf{p}(b') \times \mathbf{p}(a_1 | b') \times \mathbf{p}(a_2 | b')}{\sum_{b \in B} (\mathbf{p}(b) \times \mathbf{p}(a_1 | b) \times \mathbf{p}(a_2 | b))} \times \mathbf{p}(d | b') \quad (\text{Part 1 of this theorem}) \\
 & = \sum_{b' \in B} \frac{\mathbf{p}(b') \times \mathbf{p}(a_2 | b') \times \mathbf{p}(a_1 | b')}{\sum_{b \in B} (\mathbf{p}(b) \times \mathbf{p}(a_2 | b) \times \mathbf{p}(a_1 | b))} \times \mathbf{p}(d | b') \\
 & \approx \sum_{b \in B} ((\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(b) \times \mathbf{p}(d | b)) \\
 & = (\mathbf{p} * a_2 * a_1)(d)
 \end{aligned}$$

□

Proof of Observation 8

$$\begin{aligned}
 (\mathbf{p}[\ast_{\delta} a]^n)(b') &\approx \frac{\mathbf{p}(b') \times (\mathbf{p}(a \mid b'))^n}{\sum_{b \in B} (\mathbf{p}(b) \times (\mathbf{p}(a \mid b))^n)} && \text{(From Theorem 1, part 1)} \\
 &= \frac{\mathbf{p}(b')}{\sum_{b \in B} \left(\mathbf{p}(b) \times \left(\frac{\mathbf{p}(a \mid b)}{\mathbf{p}(a \mid b')} \right)^n \right)},
 \end{aligned}$$

which converges to 1 with increasing n . $\square$

Proof of Observation 9 The following example serves to prove both parts of the observation: Let $B = \{b_1, b_2, b_3\}$, $\mathbf{p}(b_1) = \mathbf{p}(b_2) = 0.5 - \epsilon$, $\mathbf{p}(b_3) = 2\epsilon$, $\mathbf{p}(a_1 \mid b_1) = 1$, $\mathbf{p}(a_1 \mid b_2) = 0$, $\mathbf{p}(a_1 \mid b_3) = 1$, $\mathbf{p}(a_2 \mid b_1) = 0$, $\mathbf{p}(a_2 \mid b_2) = 1$, and $\mathbf{p}(a_2 \mid b_3) = 1$. We then have:

$$\begin{aligned}
 (\mathbf{p} \ast_{\delta_1} a_1)(b_1) &= \frac{\mathbf{p}(a_1 \& b_1)}{\mathbf{p}(a_1)} + \delta_1 \left(\frac{\mathbf{p}(\neg a_1 \& b_1)}{\mathbf{p}(\neg a_1)} - \frac{\mathbf{p}(a_1 \& b_1)}{\mathbf{p}(a_1)} \right) \\
 &= \frac{0.5 - \epsilon}{0.5 + \epsilon} + \delta_1 \left(0 - \frac{0.5 - \epsilon}{0.5 + \epsilon} \right) = (1 - \delta_1) \times \frac{0.5 - \epsilon}{0.5 + \epsilon} \\
 (\mathbf{p} \ast_{\delta_1} a_1)(b_2) &= \frac{\mathbf{p}(a_1 \& b_2)}{\mathbf{p}(a_1)} + \delta_1 \left(\frac{\mathbf{p}(\neg a_1 \& b_2)}{\mathbf{p}(\neg a_1)} - \frac{\mathbf{p}(a_1 \& b_2)}{\mathbf{p}(a_1)} \right) = 0 + \delta_1 \times \left(\frac{0.5 - \epsilon}{0.5 - \epsilon} - 0 \right) = \delta_1 \\
 (\mathbf{p} \ast_{\delta_1} a_1)(b_3) &= \frac{\mathbf{p}(a_1 \& b_3)}{\mathbf{p}(a_1)} + \delta_1 \left(\frac{\mathbf{p}(\neg a_1 \& b_3)}{\mathbf{p}(\neg a_1)} - \frac{\mathbf{p}(a_1 \& b_3)}{\mathbf{p}(a_1)} \right) \\
 &= \frac{2\epsilon}{0.5 + \epsilon} + \delta_1 \left(0 - \frac{2\epsilon}{0.5 + \epsilon} \right) = (1 - \delta_1) \times \frac{2\epsilon}{0.5 + \epsilon} \\
 (\mathbf{p} \ast_{\delta_2} a_2)(b_1) &= \frac{\mathbf{p}(a_2 \& b_1)}{\mathbf{p}(a_2)} + \delta_2 \left(\frac{\mathbf{p}(\neg a_2 \& b_1)}{\mathbf{p}(\neg a_2)} - \frac{\mathbf{p}(a_2 \& b_1)}{\mathbf{p}(a_2)} \right) = 0 + \delta_2 \left(\frac{0.5 - \epsilon}{0.5 - \epsilon} - 0 \right) = \delta_2 \\
 (\mathbf{p} \ast_{\delta_2} a_2)(b_2) &= \frac{\mathbf{p}(a_2 \& b_2)}{\mathbf{p}(a_2)} + \delta_2 \left(\frac{\mathbf{p}(\neg a_2 \& b_2)}{\mathbf{p}(\neg a_2)} - \frac{\mathbf{p}(a_2 \& b_2)}{\mathbf{p}(a_2)} \right) = \frac{0.5 - \epsilon}{0.5 + \epsilon} + \delta_2 \left(0 - \frac{0.5 - \epsilon}{0.5 + \epsilon} \right) \\
 &= (1 - \delta_2) \frac{0.5 - \epsilon}{0.5 + \epsilon} \\
 (\mathbf{p} \ast_{\delta_2} a_2)(b_3) &= \frac{\mathbf{p}(a_2 \& b_3)}{\mathbf{p}(a_2)} + \delta_2 \left(\frac{\mathbf{p}(\neg a_2 \& b_3)}{\mathbf{p}(\neg a_2)} - \frac{\mathbf{p}(a_2 \& b_3)}{\mathbf{p}(a_2)} \right) = \frac{2\epsilon}{0.5 + \epsilon} + \delta_2 \left(0 - \frac{2\epsilon}{0.5 + \epsilon} \right) \\
 &= (1 - \delta_2) \left(\frac{2\epsilon}{0.5 + \epsilon} \right)
 \end{aligned}$$

It follows that:

$$\begin{aligned}
 (\mathbf{p} \ast_{\delta_1} a_1 \ast_{\delta_2} a_2)(b_1) &= \frac{(\mathbf{p} \ast_{\delta_1} a_1)(a_2 \& b_1)}{(\mathbf{p} \ast_{\delta_1} a_1)(a_2)} + \delta_2 \times \left(\frac{(\mathbf{p} \ast_{\delta_1} a_1)(\neg a_2 \& b_1)}{(\mathbf{p} \ast_{\delta_1} a_1)(\neg a_2)} - \frac{(\mathbf{p} \ast_{\delta_1} a_1)(a_2 \& b_1)}{(\mathbf{p} \ast_{\delta_1} a_1)(a_2)} \right) \\
 &= 0 + \delta_2(1 - 0) = \delta_2. \\
 (\mathbf{p} \ast_{\delta_1} a_1 \ast_{\delta_2} a_2)(b_2) &= \frac{(\mathbf{p} \ast_{\delta_1} a_1)(a_2 \& b_2)}{(\mathbf{p} \ast_{\delta_1} a_1)(a_2)} + \delta_2 \times \left(\frac{(\mathbf{p} \ast_{\delta_1} a_1)(\neg a_2 \& b_2)}{(\mathbf{p} \ast_{\delta_1} a_1)(\neg a_2)} - \frac{(\mathbf{p} \ast_{\delta_1} a_1)(a_2 \& b_2)}{(\mathbf{p} \ast_{\delta_1} a_1)(a_2)} \right)
 \end{aligned}$$

$$\begin{aligned}
&= \frac{\delta_1}{\delta_1 + (1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}} + \delta_2 \left(0 - \frac{\delta_1}{\delta_1 + (1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}} \right) \\
&= \frac{\delta_1(1 - \delta_2)}{\delta_1 + (1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}} \\
(\mathbf{p} *_{\delta_1} a_1 *_{\delta_2} a_2)(b_3) &= \frac{(\mathbf{p} *_{\delta_1} a_1)(a_2 \& b_3)}{(\mathbf{p} *_{\delta_1} a_1)(a_2)} + \delta_2 \times \left(\frac{(\mathbf{p} *_{\delta_1} a_1)(\neg a_2 \& b_3)}{(\mathbf{p} *_{\delta_1} a_1)(\neg a_2)} - \frac{(\mathbf{p} *_{\delta_1} a_1)(a_2 \& b_3)}{(\mathbf{p} *_{\delta_1} a_1)(a_2)} \right) \\
&= \frac{(1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}}{\delta_1 + (1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}} + \delta_2 \left(0 - \frac{(1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}}{\delta_1 + (1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}} \right) \\
&= (1 - \delta_2) \left(\frac{(1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}}{\delta_1 + (1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}} \right) \\
(\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(b_1) &= \frac{(\mathbf{p} *_{\delta_2} a_2)(a_1 \& b_1)}{(\mathbf{p} *_{\delta_2} a_2)(a_1)} + \delta_1 \times \left(\frac{(\mathbf{p} *_{\delta_2} a_2)(\neg a_1 \& b_1)}{(\mathbf{p} *_{\delta_2} a_2)(\neg a_1)} - \frac{(\mathbf{p} *_{\delta_2} a_2)(a_1 \& b_1)}{(\mathbf{p} *_{\delta_2} a_2)(a_1)} \right) \\
&= \frac{\delta_2}{\delta_2 + (1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}} + \delta_1 \left(0 - \frac{\delta_2}{\delta_2 + (1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}} \right) \\
&= \frac{\delta_2(1 - \delta_1)}{\delta_2 + (1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}} \\
(\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(b_2) &= \frac{(\mathbf{p} *_{\delta_2} a_2)(a_1 \& b_2)}{(\mathbf{p} *_{\delta_2} a_2)(a_1)} + \delta_1 \times \left(\frac{(\mathbf{p} *_{\delta_2} a_2)(\neg a_1 \& b_2)}{(\mathbf{p} *_{\delta_2} a_2)(\neg a_1)} - \frac{(\mathbf{p} *_{\delta_2} a_2)(a_1 \& b_2)}{(\mathbf{p} *_{\delta_2} a_2)(a_1)} \right) \\
&= 0 + \delta_1 = \delta_1 \\
(\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(b_3) &= \frac{(\mathbf{p} *_{\delta_2} a_2)(a_1 \& b_3)}{(\mathbf{p} *_{\delta_2} a_2)(a_1)} + \delta_1 \times \left(\frac{(\mathbf{p} *_{\delta_2} a_2)(\neg a_1 \& b_3)}{(\mathbf{p} *_{\delta_2} a_2)(\neg a_1)} - \frac{(\mathbf{p} *_{\delta_2} a_2)(a_1 \& b_3)}{(\mathbf{p} *_{\delta_2} a_2)(a_1)} \right) \\
&= \frac{(1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}}{\delta_2 + (1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}} + \delta_1 \left(0 - \frac{(1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}}{\delta_2 + (1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}} \right) \\
&= \frac{(1 - \delta_1)(1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}}{\delta_2 + (1 - \delta_2) \frac{2\epsilon}{0.5 + \epsilon}}
\end{aligned}$$

For an example, let $\delta_1 = \delta_2 = \epsilon$. It then follows that:

$$\begin{aligned}
(\mathbf{p} *_{\delta_1} a_1 *_{\delta_2} a_2)(b_1) &\approx 0 \\
(\mathbf{p} *_{\delta_1} a_1 *_{\delta_2} a_2)(b_2) &= \frac{\delta_1(1 - \delta_2)}{\delta_1 + (1 - \delta_1) \frac{2\epsilon}{0.5 + \epsilon}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{\epsilon(1-\epsilon)}{\epsilon + (1-\epsilon)\frac{2\epsilon}{0.5+\epsilon}} \\
&= \frac{\epsilon(1-\epsilon)}{\frac{\epsilon(0.5+\epsilon) + (1-\epsilon)2\epsilon}{0.5+\epsilon}} \\
&= \frac{\epsilon(1-\epsilon)(0.5+\epsilon)}{\epsilon(0.5+\epsilon) + (1-\epsilon)2\epsilon} \\
&= \frac{(1-\epsilon)(0.5+\epsilon)}{(0.5+\epsilon) + 2(1-\epsilon)} \approx \frac{1}{5} \\
&(\mathbf{p} *_{\delta_1} a_1 *_{\delta_2} a_2)(b_3) \approx (1-\delta_2) \left(\frac{(1-\delta_1)\frac{2\epsilon}{0.5+\epsilon}}{\delta_1 + (1-\delta_1)\frac{2\epsilon}{0.5+\epsilon}} \right) \\
&= (1-\epsilon)^2 \left(\frac{\frac{2\epsilon}{0.5+\epsilon}}{\frac{\epsilon(0.5+\epsilon) + (1-\epsilon)2\epsilon}{0.5+\epsilon}} \right) \approx \frac{4}{5} \\
&(\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(b_1) = \frac{\delta_2(1-\delta_1)}{\delta_2 + (1-\delta_2)\frac{2\epsilon}{0.5+\epsilon}} \\
&= \frac{\epsilon(1-\epsilon)}{\frac{\epsilon(0.5+\epsilon) + 2\epsilon(1-\epsilon)}{0.5+\epsilon}} = \frac{\epsilon(1-\epsilon)(0.5+\epsilon)}{\epsilon(2.5-\epsilon)} = \frac{(1-\epsilon)(0.5+\epsilon)}{2.5-\epsilon} \approx \frac{1}{5} \\
&(\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(b_2) = \delta_1 \approx 0 \\
&(\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(b_3) = \frac{(1-\delta_1)(1-\delta_2)\frac{2\epsilon}{0.5+\epsilon}}{\delta_2 + (1-\delta_2)\frac{2\epsilon}{0.5+\epsilon}} \\
&= \frac{\frac{2\epsilon(1-\epsilon)^2}{0.5+\epsilon}}{\frac{\epsilon(0.5+\epsilon) + 2\epsilon(1-\epsilon)}{0.5+\epsilon}} \\
&= \frac{2(1-\epsilon)^2}{2.5-\epsilon} \approx \frac{4}{5}
\end{aligned}$$

It follows that $(\mathbf{p} *_{\delta_1} a_1 *_{\delta_2} a_2)(a_1) \approx 0.8$ and $(\mathbf{p} *_{\delta_2} a_2 *_{\delta_1} a_1)(a_1) \approx 1$. $\square$

Proof of Observation 10 It is sufficient to prove that (1) If $a \in \llbracket \mathbf{p} \rrbracket$ and $a \vdash d$, then $d \in \llbracket \mathbf{p} \rrbracket$, and (2) If $a_1 \in \llbracket \mathbf{p} \rrbracket$ and $a_2 \in \llbracket \mathbf{p} \rrbracket$, then $a_1 \& a_2 \in \llbracket \mathbf{p} \rrbracket$.

For (1), let $a \in \llbracket \mathbf{p} \rrbracket$. If $a \vdash d$, then d is equivalent with $a \vee (d \& \neg a)$, and it follows from the third Kolmogorov axiom that $\mathbf{p}(a) \leq \mathbf{p}(a \vee (d \& \neg a)) = \mathbf{p}(d)$, hence $\text{st}(\mathbf{p}(d)) = 1$.

For (2), let $p(a_1) = 1 - \delta_1$ and $p(a_2) = 1 - \delta_2$. Then $p(\neg a_1) = \delta_1$ and $p(\neg a_2) = \delta_2$, and:

$$p(a_1 \& a_2) = 1 - p(\neg(a_1 \& a_2))$$

$$p(a_1 \& a_2) = 1 - p(\neg a_1 \vee \neg a_2)$$

$$p(a_1 \& a_2) \geq 1 - (p(\neg a_1) + p(\neg a_2)) = 1 - (\delta_1 + \delta_2) \quad (\text{Since } p(\neg a_1 \vee \neg a_2) \leq p(\neg a_1) + p(\neg a_2))$$

$$a_1 \& a_2 \in \llbracket p \rrbracket \quad (\text{Since } st(\delta_1 + \delta_2) = 0)$$

□

Proof of Observation 11 Part 1: From Observations 1 and 10.

Part 2: We will prove this part in its reverse form:

$$d \notin \text{Cn}(\llbracket p \rrbracket \cup \{a\})$$

$$a \rightarrow d \notin \llbracket p \rrbracket$$

$$p(a \rightarrow d) \not\approx 1$$

$$p(a \& \neg d) \not\approx 0$$

For some $b' \in B$: $p(a \& \neg d \& b') \not\approx 0$

$p(a \& b') \not\approx 0$, $p(\neg d \& b') \not\approx 0$, $p(a) \not\approx 0$, and $p(b') \not\approx 0$.

$$\frac{p(a \& b') \times p(\neg d \& b')}{p(a) \times p(b')} \not\approx 0$$

$$\sum_{b \in B} \frac{p(a \& b) \times p(\neg d \& b)}{p(a) \times p(b)} \not\approx 0$$

$$(p *_{\delta} a)(\neg d) \not\approx 0$$

$$(p *_{\delta} a)(d) \not\approx 1$$

$$d \notin \llbracket p *_{\delta} a \rrbracket$$

Part 3: Let $B = \{b_1, b_2\}$, $p(b_1) = p(b_2) = 0.5$, $p(a \mid b_1) = 1$, and $p(a \mid b_2) = 0.5$. Then $p(\neg a) = 0.25$, thus $\neg a \notin \llbracket p \rrbracket$. Furthermore:

$$\begin{aligned} (p *_{\delta} a)(a) &\approx \frac{(p(a \& b_1))^2}{p(a) \times p(b_1)} + \frac{(p(a \& b_2))^2}{p(a) \times p(b_2)} \\ &= \frac{0.5^2}{0.75 \times 0.5} + \frac{0.25^2}{0.75 \times 0.5} = \frac{5}{6} \end{aligned}$$

Thus, $a \notin \llbracket p *_{\delta} a \rrbracket$.

Part 4: Let $0 \ll r \ll 1$ and let B be such that $p(a \mid b) \leq r$ for all $b \in B$. It follows that $(p *_{\delta} a)(d) \leq r$ for all d and δ .

Parts 5-6: These results follow directly from the definitions. □

Proof of Observation 12 Part 1: Let $B = \{b_1, b_2\}$, $p(b_1) = 1 - \epsilon$, $p(b_2) = \epsilon$, $p(a \mid b_1) = p(c \mid b_1) = 0$, $p(a \mid b_2) = 0.5$, and $p(c \mid b_2) = 1$.

We then have $a \notin \llbracket p *_\delta a \rrbracket$ and $c \notin \llbracket p \rrbracket$. However:

$$\begin{aligned} (p *_\delta a)(c) &= (1 - \delta) \frac{p(a \& b_2) \times p(c \& b_2)}{p(a) \times p(b_2)} + \delta \frac{p(\neg a \& b_2) \times p(c \& b_2)}{p(\neg a) \times p(b_2)} \\ &= (1 - \delta) \frac{0.5\epsilon \times \epsilon}{0.5\epsilon \times \epsilon} + \delta \frac{0.5\epsilon \times \epsilon}{(1 - 0.5\epsilon) \times \epsilon} \approx 1 \end{aligned}$$

and thus $c \in \llbracket p *_\delta a \rrbracket$.

Part 2: Let $B = \{b_1, b_2\}$, $p(b_1) = p(b_2) = p(a \mid b_1) = 0.5$, $p(d \mid b_1) = 0$, and $p(a \mid b_2) = p(d \mid b_2) = 1$. It follows that for all δ :

$$\begin{aligned} (p *_\delta a)(a) &\approx \frac{(p(a \& b_1))^2}{p(a) \times p(b_1)} + \frac{(p(a \& b_2))^2}{p(a) \times p(b_2)} = \frac{0.25^2}{0.75 \times 0.5} + \frac{0.5^2}{0.75 \times 0.5} = \frac{5}{6} \\ (p *_\delta d)(a) &\approx \frac{p(d \& b_1) \times p(a \& b_1)}{p(d) \times p(b_1)} + \frac{p(d \& b_2) \times p(a \& b_2)}{p(d) \times p(b_2)} = 0 + \frac{0.5 \times 0.5}{0.5 \times 0.5} = 1 \end{aligned}$$

□

Proof of Observation 13 Let $\neg a \notin \llbracket p \rrbracket$ and $d \in \llbracket p \rrbracket$. We are going to show that $d \in \llbracket p *_\delta a \rrbracket$.

First step: First, we are going to show that for all $b' \in B$, if $p(b' \mid a) \not\approx 0$, then $p(b') \not\approx 0$. Let $p(b' \mid a) \not\approx 0$, i.e. $p(a \& b')/p(a) \not\approx 0$. It follows from $\neg a \notin \llbracket p \rrbracket$ that $p(a) \not\approx 0$, thus $p(a \& b') \not\approx 0$, thus $p(b') \not\approx 0$.

Second step: We are going to show that for all $b' \in B$, if $p(b') \not\approx 0$, then $p(d \mid b') \approx 1$. Let $p(b') \not\approx 0$. It follows from $d \in \llbracket p \rrbracket$ that $\sum_{b \in B} (p(b) \times p(d \mid b)) \approx 1$. We also have $\sum_{b \in B} p(b) = 1$ and $p(d \mid b) \leq 1$ for all $b \in B$, from which the desired conclusion follows.

Third step:

$$\begin{aligned} (p *_\delta a)(d) &\approx \sum_{b \in B} (p(b \mid a) \times p(d \mid b)) \\ &\approx \sum_{b \in B} p(b \mid a) && \text{(Steps 1 and 2)} \\ &= 1 \end{aligned}$$

□

Proof of Observation 14 Part 1: Let $B = \{b_1, b_2\}$, $p(b_1) = p(b_2) = 0.5$, $p(a_1 \& a_2 \& d \mid b_1) = 1$, and $p(a_1 \neg \& a_2 \& \neg d \mid b_2) = p(\neg a_1 \& a_2 \& \neg d \mid b_2) = 0.5$. Then:

$$\begin{aligned} (p *_\delta (a_1 \& a_2))(d) &= \\ &\approx \frac{p(a_1 \& a_2 \& b_1) \times p(d \& b_1)}{p(a_1 \& a_2) \times p(b_1)} + \frac{p(a_1 \& a_2 \& b_2) \times p(d \& b_2)}{p(a_1 \& a_2) \times p(b_2)} \\ &= \frac{0.5^2}{0.5^2} + 0 = 1. \end{aligned}$$

Thus, $d \in \llbracket p *_{\delta} (a_1 \& a_2) \rrbracket$. Furthermore:

$$\begin{aligned} (p *_{\delta'} a_1)(a_2 \rightarrow d) &\approx \\ &\approx \frac{p(a_1 \& b_1) \times p((a_2 \rightarrow d) \& b_1)}{p(a_1) \times p(b_1)} + \frac{p(a_1 \& b_2) \times p((a_2 \rightarrow d) \& b_2)}{p(a_1) \times p(b_2)} \\ &= \frac{0.5 \times 0.5}{0.75 \times 0.5} + \frac{0.25 \times 0.25}{0.75 \times 0.5} = \frac{2}{3} + \frac{1}{6} = \frac{5}{6} \end{aligned}$$

Thus, $a_2 \rightarrow d \notin \llbracket p *_{\delta'} a_1 \rrbracket$ and consequently $d \notin \text{Cn}(\llbracket p *_{\delta'} a_1 \rrbracket \cup \{a_2\})$.

Part 2: Let $B = \{b_1, b_2\}$, $p(b_1) = p(b_2) = 0.5$ and $p(a_1 \& a_2 \& b_1) = p(\neg a_1 \& \neg a_2 \& b_1) = p(a_1 \& a_2 \& b_2) = p(\neg a_1 \& \neg a_2 \& b_2) = 0.25$. Then:

$$\begin{aligned} (p *_{\delta} a_1)(\neg a_2) &\approx \\ &\approx \frac{p(a_1 \& b_1) \times p(\neg a_2 \& b_1)}{p(a_1) \times p(b_1)} + \frac{p(a_1 \& b_2) \times p(\neg a_2 \& b_2)}{p(a_1) \times p(b_2)} \\ &= \frac{0.25 \times 0.25}{0.5 \times 0.5} + \frac{0.25 \times 0.25}{0.5 \times 0.5} = 0.5 \end{aligned}$$

and:

$$\begin{aligned} (p *_{\delta'} (a_1 \& a_2))(a_2) &\approx \\ &\approx \frac{p(a_1 \& a_2 \& b_1) \times p(a_2 \& b_1)}{p(a_1 \& a_2) \times p(b_1)} + \frac{p(a_1 \& a_2 \& b_2) \times p(a_2 \& b_2)}{p(a_1 \& a_2) \times p(b_2)} \\ &= \frac{0.25 \times 0.25}{0.5 \times 0.5} + \frac{0.25 \times 0.25}{0.5 \times 0.5} = 0.5 \end{aligned}$$

Thus $\neg a_2 \notin \llbracket p *_{\delta} a_1 \rrbracket$ and $a_2 \notin \llbracket p *_{\delta'} (a_1 \& a_2) \rrbracket$. □

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D. (1985). On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50, 510–530.
2. Benci, V., Horsten, L., Wenmackers, S. (2013). Non-Archimedean probability. *Milan Journal of Mathematics*, 81, 121–151.
3. Benci, V., Horsten, L., Wenmackers, S. (2018). Infinitesimal probabilities. *British Journal for the Philosophy of Science*, 69(2), 509–552.
4. Cantwell, J. (1997). On the logic of small changes in hypertheories. *Theoria*, 63, 54–89.
5. Fermé, E., & Hansson, S.O. (2018). *Belief change. Introduction and overview*. Springer.
6. Foley, R. (1992). The epistemology of belief and the epistemology of degrees of belief. *American Philosophical Quarterly*, 29(2), 111–124.
7. Gaifman, H. (1988). A theory of higher order probabilities. In Skyrms, B., & Harper, W.L. (Eds.) *Causation, chance and credence. Proceedings of the Irvine conference on probability and causation*, (Vol. 1 pp. 191–219). Dordrecht: Kluwer.
8. Gärdenfors, P. (1986). Belief revisions and the Ramsey test for conditionals. *Philosophical Review*, 95(1), 81–93.
9. Gärdenfors, P. (1988). *Knowledge in flux. Modeling the dynamics of epistemic states*. Cambridge: The MIT Press.
10. Hájek, A. (2003). What conditional probability could not be. *Synthese*, 137(3), 273–323.
11. Hansson, S.O. (1994). Taking belief bases seriously. In Prawitz, D., & Westerståhl, D. (Eds.) *Logic and philosophy of science* (pp. 13–28). Uppsala: Kluwer.

12. Hansson, S.O. (1996). A test battery for rational database updating. *Artificial Intelligence*, 82, 341–352.
13. Hansson, S.O. (1996). What is philosophy of risk? *Theoria*, 62, 169–186.
14. Hansson, S.O. (1997). Semi-revision. *Journal of Applied Non-Classical Logic*, 7, 151–175.
15. Hansson, S.O. (1999). *A textbook of belief dynamics*. Dordrecht: Kluwer.
16. Hansson, S.O. (1999). A survey of non-prioritized belief revision. *Erkenntnis*, 50, 413–427.
17. Hansson, S.O. (2008). Do we need second-order probabilities? *Dialectica*, 62, 525–533.
18. Hansson, S.O. (2017). *Descriptor revision. Belief change through direct choice*. Springer.
19. Hansson, S.O. (2018). Review of Hannes Leitgeb, *The stability of belief. How Rational belief Coheres with probability*. *Journal of Philosophy*, 115(5), 276–280.
20. Hansson, S.O. (2018). Politique du risque et l'intégrité de la science. In de Guay, A. (Ed.) *Risque et Expertise. Les Conférences Pierre Duhem* (pp. 57–112). Besançon: Presses universitaires de Franche-Comté.
21. Hansson, S.O., Fermé, E., Cantwell, J., Falappa, M. (2001). Credibility-limited revision. *Journal of Symbolic Logic*, 66, 1581–1596.
22. Hedden, B. (2019). Hindsight bias is not a bias. *Analysis*, 79(1), 43–52.
23. Hrbacek, K. (2007). Stratified analysis? In van den Berg, I., & Neves, V. (Eds.) *The strength of nonstandard analysis* (pp. 47–63). Wien: Springer.
24. Jeffrey, R.C. (1983). *The logic of decision*, 2nd edn. Chicago: University of Chicago Press.
25. Keisler, H.J. (1986). *Elementary calculus. An infinitesimal approach*, 2nd edn. Boston: Prindle, Weber & Schmidt.
26. Keller, A.M., & Wilkins, M.W. (1985). On the use of an extended relational model to handle changing incomplete information. *IEEE Transactions on Software Engineering*, 11(7), 620–633.
27. Kolmogorov, A.N. ([1933] 1956). *Foundations of the theory of probability*, second English edition. New York: Chelsea.
28. Konieczny, S., Grespan, M.M., Pérez, R.P. (2010). Taxonomy of improvement operators and the problem of minimal change. In 161–170 (Ed.) *Proceedings of the 12th international conference on principles of knowledge representation and reasoning (KR 2010)*. Menlo Park: AAAI Press.
29. Konieczny, S., & Pérez, R. P. (2008). Improvement operators. In *Proceedings of the 11th international conference on principles of knowledge representation and reasoning (KR08)* (pp. 177–186). Menlo Park: AAAI Press.
30. Kyburg, H. (1961). *Probability and the logic of rational belief*. Middletown: Wesleyan University Press.
31. Leitgeb, H. (2017). *The stability of belief: how rational belief coheres with probability*. Oxford: Oxford University Press.
32. Lewis, D. ([1980] 1987). A subjectivist's guide to objective chance. In Lewis, D. (Ed.) *Philosophical papers*, (Vol. II pp. 83–132). Oxford: Oxford University Press. First published in Richard C. Jeffrey, ed., (1980) *Studies in Inductive Logic and Probability*, volume II, pp. 263–293. Berkeley University of California Press.
33. Lyon, A. (2016). Kolmogorov's axiomatization and its discontents. In Hájek, A., & Hitchcock, C. (Eds.) *The Oxford handbook of probability and philosophy* (pp. 155–166). Oxford: Oxford University Press.
34. Makinson, D. (1965). The paradox of the preface. *Analysis*, 25, 205–207.
35. Rott, H. (2001). *Change choice and inference: a study of belief revision and nonmonotonic reasoning*. Oxford: Clarendon.
36. Rott, H., & Hansson, S.O. (2014). Safe contraction revisited. In Hansson, S.O. (Ed.) *David Makinson 829 on classical methods for non-classical problems* (pp. 35–70). Springer.
37. Schurz, G. (2019). Impossibility results for rational belief. *Noûs*, 53(1), 134–159.
38. Seidenfeld, T. (2001). Remarks on the theory of conditional probability: some issues of finite versus countable additivity. In Hendricks, V.F., Pedersen, S.A., Jørgensen, K.F. (Eds.) *Probability theory: philosophy, recent history and relations to science* (pp. 167–178). Dordrecht: Kluwer.
39. Seidenfeld, T., Schervish, M.J., Kadane, J.B. (2001). Improper regular conditional distributions. *Annals of Probability*, 29(4), 1612–1624.
40. Shimony, A. (1955). Coherence and the axioms of confirmation. *Journal of Symbolic Logic*, 20(1), 1–28.
41. Skyrms, B. (1980). *Causal necessity: a pragmatic investigation of the necessity of laws*. New Haven: Yale University Press.

42. Skyrms, B. (1980). Higher order degrees of belief. In Mellor, D.H. (Ed.) *Prospects for pragmatism. Essays in memory of F.P Ramsey* (pp. 109–137). Cambridge: Cambridge University Press.
43. Skyrms, B. (1983). Three ways to give a probability assignment a memory. In Earman, J. (Ed.) *Testing scientific theories* (pp. 157–161). Minneapolis: University of Minnesota Press.
44. Wenmackers, S. (2013). Ultralarge lotteries: analyzing the lottery paradox using non-standard analysis. *Journal of Applied Logic*, 11, 452–467.
45. Wenmackers, S., & Horsten, L. (2013). Fair infinite lotteries. *Synthese*, 190, 37–61.
46. Wenmackers, S., & Romeijn, J.-W. (2016). New theory about old evidence. *Synthese*, 193(4), 1225–1250.
47. Zhuang, Z., Delgrande, J., Nayak, A., Sattar, A. (2017). A unifying framework for probabilistic belief revision. In *Proceedings of the 26th international joint conference on artificial intelligence (IJCAI-17)* (pp. 1370–1376). <https://www.ijcai.org/proceedings/2017/0190.pdf>. Accessed June 8, 2019.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.