THE JOURNAL OF PHILOSOPHY
VOLUME XCVIII, NO. 11, NOVEMBER 2001

BAYESIAN EPISTEMOLOGY AND EPISTEMIC CONDITIONALS:
ON THE STATUS OF THE EXPORT-IMPORT LAWS*

The notion of probability occupies a central role in contemporary epistemology and cognitive science. Nevertheless, the classical notion of probability is hard to reconcile with the central notions postulated by the epistemological tradition. ¹ Bayesian epistemologists have presented three main types of responses to this problem. First, there are eliminative strategies: if notions like belief cannot be probabilistically articulated, they have no theoretical use

* I am grateful to Rohit Parikh, Teddy Seidenfeld, and Bas van Fraassen for helpful comments and suggestions. This work was supported in part by a grant from the National Science Foundation.

¹ A classical reference in this area is Henry Kyburg, Probability and the Logic of Rational Belief (Middletown: Wesleyan, 1961). One of the paradoxes presented in Kyburg’s book, the so-called lottery paradox, is now well known in the literature and will be considered here in detail. The moral of the paradox can be summarized as follows: it shows that, if one accepts as epistemically rational to believe propositions just in case it is rational to have a high degree of confidence in them, then one has to accept as well that it is epistemically rational to be in weakly inconsistent epistemic positions, where currently held propositions do not have a nonempty intersection. Epistemologists who insist on logical closure and consistency as requirements for rational belief reject the possibility of this consequence, and, therefore, they also reject modeling qualitative doxastic states in terms of threshold value for degree-of-confidence functions. Of course, one’s philosopher modus ponens can be easily transmuted into another’s philosopher modus tollens. So, epistemologists, like Kyburg himself, have argued that the scenario of the lottery gives us good reason to adopt the threshold value model and to reject (the full strength of) logical closure as a doxastic ideal of rationality. The latter solution is compatible with a probabilistic stance in epistemology (where probability is the only primitive), but it delivers a notion of belief that does not connect smoothly with the received view in the epistemological tradition. The former solution is incompatible with probabilism due to the fact that it conflicts with the adoption of acceptance rules in terms of high probability, or even probability one (or probability infinitesimally close to one).
and, consequently, they have to be eliminated from a mature epistemology. Second, there are pluralist strategies, which recommend adopting both belief and probability as irreducible primitive notions. Finally, there is the claim that a unified Bayesian epistemology can be articulated around the sole postulation of a primitive notion of conditional probability. I shall focus here on two aspects of this third probabilistic strategy. First, I shall define both belief and belief change from conditional probability—following an idea first proposed by Bas van Fraassen.\(^2\) Second, I shall argue that the Adams hypothesis (on how to accept conditionals\(^3\)) can be better articulated and extended in this setting. Finally, I shall discuss both the internal commitments and the range of applicability of the probabilistic view of belief and conditionals which thus arises.

I. INTRODUCTION

van Fraassen\(^4\) has tried to find a common ground between probabilism and the received view in epistemology: “What I hope for is some reconciliation of the diverse intuitions of Bayesians and traditionalists, within a rather liberal probabilism” (ibid., p. 170). In a more technical article, van Fraassen both reconsidered the motivations for pursuing this form of unified probabilism and proposed various concrete manners of formulating it:

Personal or subjective probability entered epistemology as a cure for certain perceived inadequacies in the traditional notion of belief. But there are severe strains in the relationship between probability and belief. They seem too intimately related to exist as separate but equal; yet if either is taken as the more basic, the other may suffer. [...] I would like to propose a single unified account, which takes conditional personal probability as basic (op. cit., p. 349).

Epistemologically, the idea is to appeal to a third notion (besides qualitative belief and subjective grading), namely, the notion of supposition, and to encode it via a pre-Kolmogorovian notion of conditional probability:

There is a third aspect of opinion, besides belief and subjective grading, namely supposition. Much of our opinion can be elicited only by asking us to suppose something, which we may or may not believe. Here


supposition will be a central ingredient of an unified probabilistic picture (op. cit., p. 351).

One of my main goals here is to articulate, extend, and explore the limits of this unified form of probabilism. I shall consider not only how to define qualitative belief and expectations in a paradox-free manner, but also how these doxastic notions change when new items are supposed or learned. I shall characterize the correspondent notion of probability-based belief change, and compare it with nonprobabilistic accounts of belief change recently developed in the literature.

It is important to realize that the noneliminative probabilism just considered has been postulated not only in order to define a unified epistemological view, but also in order to reform and improve the traditional notion of probability itself. The following example can help to clarify this point.

Consider a map of the Western Hemisphere and focus your attention on any two points on the map. Say that you select a point \( N \) representing New York City and that you choose a second point \( B \) representing Baltimore. In addition, call \( W \) the surface representing the Western Hemisphere, \( S \) the surface representing the state of New York, and \( M \) the surface representing Maryland. Now pick a point \( P \) at random in \( W \). The (unconditional) probability that \( P \) is either \( N \) or \( B \) is zero, since \( P \) has been chosen from an infinite number of points. In spite of this fact, it seems that one would refuse to say that the probability that the arbitrarily chosen point is \( N \) is undefined, given that the point is either \( N \) or \( B \). Moreover, it seems rather natural to say that such probability is exactly \( 1/2 \).

Two main strategies have been implemented to solve this problem. On the one hand, one can take infinitesimal numbers at face value. One can say that the proposition that the randomly chosen point \( P \) is \( N \) carries an infinitesimal probability \( i \). Such probability is a nonstandard number greater than zero but strictly smaller than any standard positive number. By the same token, the probability that the

---

5 This example is inspired by a similar example initially proposed by A. N. Kolmogoroff, and by a variant of it recently proposed by Adams in his A Primer of Probability Logic (Stanford: CSLI, 1988), pp. 252-53.

6 There are, to be sure, alternative approaches to the ones I am about to present, but they are not available to epistemologists of probabilistic persuasion. I shall focus here on solutions that can be formulated in the context of a probabilistic epistemology. Section 11, on the other hand, begins to articulate the distinctions that separate relevant varieties of Bayesian epistemology. With the help of these distinctions, I shall also sketch the gist of alternative approaches to the ones considered in this section.
point is either $N$ or $B$ carries $2i$ value. And the conditional probability that $P$ is $N$ given that $P$ is either $N$ or $B$ would then carry the value $i/2i$. This representation of the problem begins to approximate intuition. Now, the controversial probability of $P$ being $N$ given that $P$ is either $N$ or $B$ is defined. In addition, infinitesimal numbers obey arithmetical laws that allow us to cancel the infinitesimal $i$ in the quotient $i/2i$. And this yields the desired result $1/2$.

At first sight, the move might appear as a rather natural response to the problem of how to condition on events of (standard) measure zero. Nevertheless, the strategy causes as many problems as it seems to solve. Many philosophers and mathematicians have for centuries considered infinitesimal numbers with suspicion. As a result, infinitesimals have been often replaced in applications by limits of series of standard numbers. At least this was so until Abraham Robinson made them respectable again in 1966 with his important *Non-standard Analysis*. Today, the theory of infinitesimal probability is well understood, and many philosophers have officially adopted it as an improvement of standard Bayesian theory; for example, Brian Skyrms, David Lewis, and Vann McGee have recently argued in favor of this view.

Several well-known properties of ordered algebraic fields are violated in this new setting. A salient one is the so-called *Archimedean axiom* establishing that for any positive number $r$, no matter how small, there is an integer $n$ such that $nr > 1$. A non-Archimedean field will contain at least an infinitesimal element $\varepsilon$ with the property that $n\varepsilon < 1$ for every positive integer $n$, even though $\varepsilon$ is itself positive. A nice property of non-Archimedean subjective probability is that it obeys all the Kolmogorovian axioms extended to nonstandard values. Nevertheless, many foundational problems remain open; for example, infinitesimals are not measurable.

The most salient alternative to the use of infinitesimal numbers is based upon the work of Karl Popper and Alfred Renyi. In standard probability theory, absolute probability is adopted as a primitive, and characterized axiomatically via Kolmogoroff's axioms. Conditional

---

probability is then defined in terms of absolute probability. Both Popper and Renyi proposed to study a theory where this idea is completely inverted.\textsuperscript{10} Conditional probability is adopted as a primitive, and absolute probability is defined as conditional probability given the truth.\textsuperscript{11} Then the notion of conditional probability is axiomatically characterized in such a way that the probability of a proposition $A$ given $B$ can be meaningfully defined even when the unconditional probability of $B$ is zero. van Fraassen appealed to a variant of the Popper-Renyi strategy in order to construct a unified version of probabilism.

The Popper-Renyi view has its own foundational problems, some of which are considered in sections VIII and IX below. Nevertheless, the Popper-Renyi’s two-place probability functions can be interestingly applied when the underlying probabilistic space is at most of countable size. In addition, a result recently proved by McGee (op. cit.) shows that the standard values of infinitesimal probability functions are representable as two-place probability functions, and that every two-place function is representable in terms of the standard real values of some infinitesimal probability function. McGee’s arguments show that there is a robust and hidden unity linking the two strategies used to define generalized conditional probability. Both strategies capture aspects of the intuitive notion of probability, which the post-Kolmogorovian account seemed to have failed to capture. The trade-off is that the new strategies lack solid foundations in measure theory. My goal here is not to close this foundational gap, but to examine the claims that have been made in favor of the use of two-place functions. As we just

\textsuperscript{10} The underlying motivations of Popper and Renyi were different; in addition, the theories they proposed differ in important aspects that I cannot review here in detail.

\textsuperscript{11} The underlying idea here is that there are no degrees of belief defined in a vacuum. All belief is conditional given some body of background knowledge. “Unconditional” degree of belief is, therefore, an artificial limit case where the judgments are made against no background of empirical hypotheses. But, of course, the probabilistic “status quo” includes all a priori logical truths. This is a consequence of the fact that probabilistic representations of attitudes are affected by the so-called problem of logical omniscience. All probabilistic agents are perfect logicians, or, to put it in a different manner, the minimal admissible background knowledge a probabilistic agent can admit is one where he is certain of all logical laws. Philosophers have different views as to the reasonability of this feature of probabilistic representations. I want to remark that logical omniscience will not be put into question here. It should be said in passing also that nothing precludes saying that the probability of $B$ given $A$ is the probability that $B$ is true, when $A$ is a tautology—see D. Edgington, “Lowe on Conditional Probability,” \textit{Mind}, civ, 420 (October 1996): 617-30.
saw, one important step toward the formulation of a unified probabilism is to produce a paradox-free account of belief and belief change. Another important challenge for the formulation of a unified probabilism is to produce an adequate account of how we accept and use conditional sentences.

Frank Ramsey articulated the view, now common, according to which the conditional probability of $B$ given $A$ should be viewed as the degree of belief of $B$ under the supposition that $A$ (ibid., p. 76). He was also responsible for a very influential view of conditionals, which he understood as "cognitive carriers." According to Ramsey's view, conditional sentences lack truth conditions, but they have precise acceptability conditions, which can be formulated inside a unified probabilistic epistemology. The idea is that to evaluate a conditional we add its antecedent hypothetically to our stock of knowledge, and then assess the consequent on that basis. As Ramsey puts it, in that case we "are fixing our degree of belief in [the consequent of the evaluated conditional] given [the antecedent]." But, "if [the antecedent] turns out false, these degrees of belief are rendered void" (ibid., p. 154, footnote). We can appeal to the geographical example provided above in order to illustrate this point. What is the probability of the conditional, 'If the randomly chosen point $P$ coincides either with $N$ or $B$, then it lies either on the surface representing New York ($S$) or on the surface representing Maryland ($M$)'? Ernest Adams, in his own formulation of Ramsey's account, has proposed to set this probability to one by default; in this case, the default value seems to match intuition. Nevertheless, consider the following two conditionals: (1) 'If the randomly chosen point $P$ coincides either with $N$ or $B$, then it lies on the surface $S$'; (2) 'If the randomly chosen point $P$ coincides either with $N$ or $B$, then it does not coincide with $N$, and it does not coincide with $B$'.

One might argue that the probability of both conditionals is defined, and that the probability of the first conditional is $1/2$ and the

---

13 This view is clearly expressed by Ramsey in "General Propositions and Causal- ity," in Philosophical Papers, pp. 145-63.
15 Adams's classical account of conditionals appears in The Logic of Conditionals (Boston: Reidel, 1975). Nevertheless, in this book Adams avoids making a decision about what to do with probability zero antecedents. He uses assignments where every factual sentence has positive probability and where the language does not contain conditionals with inconsistent antecedents.
probability of the second is zero rather than one. McGee has recently proposed the same remedy as the one considered above in order to treat this problem; that is, he proposed an improved version of the Ramsey-Adams hypothesis (degree of acceptability of conditionals go by conditional probability) where the underlying notion of conditional probability is generalized conditional probability. Here I intend to extend and articulate further McGee’s approach. An important limitation of his proposal is that it does not cover iterated conditionals, so I shall provide models of acceptance for those conditionals. I shall also claim that this extension provides valuable insights about the nature of conditional inference induced by unified probabilism (and on the nature of unified probabilism itself). The analysis of the so-called export-import axiom (‘If \( \phi \), then if \( \psi \), then \( \chi \)’ if and only if ‘If \( \phi \) and \( \psi \), then \( \chi \)’\(^{16}\)), for example, will shed light on the unique features of the probabilistic construal of conditional inference. In fact, although I shall argue that probabilistic models of conditionals require the law, almost no rival account endorses it.\(^{17}\) This applies not only to the most salient accounts in terms of possible worlds (for example, the models offered by Lewis, Robert Stalnaker, Donald Nute, or John Pollock\(^{18}\)), but also to the so-called epistemic models in terms of qualitative notions of supposition.\(^{19}\) In addition, several articles questioning the validity of the export law for English conditionals have been recently published in this journal.\(^{20}\) I shall show, nevertheless, that export-import is a deeply entrenched commitment of unified probabilism.

It is perhaps important to see what type of argument I am offering. Rather than defend the a priori validity of the law or the fact that

\(^{16}\) The variable \( \chi \) ranges over conditional sentences eventually containing conditionals, while \( \phi \) and \( \psi \) range over conditional-free sentences.

\(^{17}\) Probabilists have offered arguments in favor of the law (usually by appealing to qualitative models). McGee, for example, offers an argument of this type in “A Counterexample to Modus Ponens,” this journal, lxxxii, 9 (September 1985: 462-71. Frank Jackson also defended the law in his Conditionals (New York: Blackwell, 1987), appendix A1, p. 133. Edgington also argued in favor of exportation and importation in her state of the art article on conditionals in Mind, cix, 414 (1995): 235-329. Adams has a more detached view, prima facie favorable to the law.


English conditionals obey it, I shall focus on showing that the law arises as a natural commitment for defenders of certain forms of unified probabilism. I shall also argue that the problem of finding interconnections between qualitative and quantitative doxastic states is intimately related to the problem of clarifying the account of conditionals induced by unified probabilism. In particular, I shall show that the validity of the export-import law itself depends on robust properties of probability-based belief change.

As van Fraassen suggests, much of our opinion (belief, belief change, and conditionals, among other notions) can be elicited from suppositions. I intend here to make this process of elicitation and reduction explicit. It is also my purpose to reveal the important epistemological constraints imposed by it. In a nutshell, two of my main interests here are: first, to show how a unified probabilism can be possible; and, second, to determine which are some of its most salient commitments.

II. MONISM AND PLURALISM IN BAYESIAN EPISTEMOLOGY

Richard Jeffrey has recently coined the term ‘radical probabilism’. I hope to delimit this view from other forms of Bayesian epistemology and to make clear what the differences (similarities) are separating (uniting) these approaches. There is a “family resemblance” linking van Fraassen’s and Jeffrey’s varieties of radical probabilism, which can be described in Jeffrey’s own words:

Radical probabilism doesn’t insists that probabilities be based on certainties; it can be probabilities all the way down, to the roots (ibid., p. 11). [...] Radical probabilism adds the ‘non-foundational’ thought that there is no bedrock of certainty underlying our probability judgments (ibid., pp. 44-45).

This form of “nonfoundationalism” is shared between van Fraassen’s and Jeffrey’s forms of radical probabilism. Unlike Bruno De Finetti or L. J. Savage, they appeal to a form of probabilistic monism according to which certainties (or full beliefs) are not used in order to determine the space over which rational agents attribute probabilities. Both theories adopt a radical stance by reducing the set of admissible epistemological primitives to just a single probabilistic concept. Any other epistemological notion is legitimate only if it can be derived or explicated in terms of this single primitive. In contrast, De Finetti has eloquently described a pluralistic strategy in Bayesian epistemology:

---

In almost all circumstances, and at all times, we find ourselves in a state of uncertainty. Uncertainty in every sense. [...] It would therefore seem natural that the customary modes of thinking, reasoning and deciding should hinge explicitly and systematically on the factor uncertainty as the conceptually pre-eminent and determinative element. The opposite happens however: there is no lack of expressions referring to uncertainty, but it seems that these expressions, by and large, are no more than verbal padding. The solid, serious, effective and essential part of arguments, on the other hand, would be the nucleus that can be brought within the language of certainty—of what is certainly true or certainly false. It is in this ambit that our faculty of reasoning is exercised, habitually, intuitively and often unconsciously (ibid., p. 24).

Immediately, De Finetti makes clear that his set of certainties contains more than mere tautologies, and that its main role is to determine a space of possibilities:

Thinking of a subset of truths as given (knowing, for instance, that certain facts are true, certain quantities have given values, or values between certain limits, certain shapes, bodies or graphs of given phenomena enjoy certain properties, and so on), we will be able to ascertain which conclusions, among those of interests, will turn out to be—on the basis of the data—either certain (certainly true), or impossible (certainly false), or else possible (ibid., p. 25).

What about probability? According to De Finetti "probability is something that can be distributed over the field of possibility":

Using a visual image, which at a later stage could be taken as an actual representation, we could say that the logic of certainty reveals to us a space in which the range of possibilities is seen in outline, whereas the logic of the probable will fill in this blank outline by considering a mass distributed upon it (ibid., p. 25).

De Finetti usually refers to this assumed body of certainties as the state of information of the agent. It is important to realize that the adoption of a pluralist stance does not preclude recognizing the cognitive primacy of the notion of conditional probability.

We have at times insisted on making clear the fact that every prevision, and, in particular, every evaluation of probability, is conditional; not only on the mentality or psychology of the individual involved, at the time in question, but also, and especially on the state of information in which he finds himself at that moment (ibid., p. 134).
Notable subscribers of this pluralistic view include I. J. Good, Levi, and Savage. Notice that full belief (certainty) and conditional probability are introduced here as two independent primitives. In a sense certainty is the true primitive. Probability can be defined only when the dual of certainty (possibility) is specified. Moreover, in this framework it is perfectly possible to study several static and dynamic properties of bodies of certainty independently of the probabilistic aspects of the model. In other words, belief change is not determined by probability change. This is the degree of independence missing when a monist stance is adopted. Even if one can specify in an unproblematic manner belief as a function of probability, belief change is now completely determined by probability change. The correspondent notion was called above probability-based belief change. Most of the work in the first part of the present discussion will be devoted to the clarification of the properties of probability-based belief change induced by unified probabilism, that is, by a noneliminative probabilistic strategy.

III. ADAMS’S HYPOTHESIS AND ACCEPTABILITY OF CONDITIONALS

Adams has proposed the following hypothesis in a series of articles and books since, at least, the mid 1960s (op. cit.).

Original Adams hypothesis: the probability of a simple conditional ‘If a, then b’ will be \( cp(B|A) = \frac{P(A \cap B)}{P(A)} \), if \( P(A) \) is nonzero, and 1 otherwise—‘A’ denotes the proposition expressed by ‘a’, and the same applies to ‘B’ and ‘b’. In most versions of Bayesian epistemology, credence in a proposition \( X \) is the probability measure of the degree of partial or full belief in \( X \). Conditional credence, on the other hand, is represented in terms of the conditional probability \( cp \). As D.H. Mellor points out in a recent article, conditional credences are not credences: \( P(A \cap B) | P(A) \) is a

---


24 Jeffrey’s radical probabilism is a paradigmatic example of an eliminative type of probabilistic epistemology.

25 See also McGee’s presentation of the Adams hypothesis in “Learning the Impossible.” As I explained above, the Adams hypothesis appears in its most complete form in “Probability and the Logic of Conditionals.”

26 I shall present more detailed conventions about notation at the end of section II.

function of two credences, not a credence in its own right. Conditional credences are, nevertheless, probabilities (they satisfy standard axioms). So, we can adopt the following reformulation of the Adams hypothesis (due to Mellor (ibid., p. 233)): the degree of acceptance of ‘If a, then b’ is determined by the conditional credence \( cp(B|A) \), as long as such conditional credence is defined. As I explained above, the default clause (for the case in which \( cp(B|A) \) is undefined) is unreasonable. A remedy is to start with a primitive notion of conditional probability, and to constrain it with suitable axioms, allowing for well-defined conditional probabilities \( P(A|B) \), where \( B \) can carry probability zero. Following recent terminological conventions, I shall call such probabilities two-place probability functions. The new test is then reformulated as follows:

**Improved Adams hypothesis**: the degree of acceptance of ‘If a, then b’ is given by \( P(B|A) \), where \( P(-|-) \) is a two-place probability function.

And, of course, one can distill a test for full acceptance of conditionals from this improved version of the Adams hypothesis. In fact, this test for acceptance will just be a limit case of the improved Adams hypothesis:

**Improved Adams hypothesis for acceptance**: ‘If a, then b’ is accepted with respect to \( P(-|-) \) if and only if \( P(B|A) = 1 \).

According to Adams, degrees of acceptance go by degrees of conditional credence. So, these degrees of acceptance are not credences on their own right, but they are probabilities. According to the improved version, degrees of acceptance (and full acceptance) are encoded as degrees of a primitively given notion of conditional probability. This new account preserves the formal features of probability in countable spaces, but it might not be proper to refer to it as a probability measure either. According to the received view in conditional probability, if \( w \) is a point in a probability space with universe \( U \), \( P(-|w) \) is not probability given an (arbitrary) event, but probability given a sigma field. A more comprehensive discussion of this point is offered in sections VIII and IX. Here I want to point out only that the adoption of the improved version of the Adams hypothesis requires a further departure from the idea that the measure of acceptance of conditionals is a probability measure. The original Adams hypothesis preserves the idea that the measure of degrees of acceptance of conditionals, even if not a credence, is a probability measure (it is a conditional credence). The adoption of the improved version
removes various problems related to the original version, but it would be better to say that the degrees of acceptance it characterizes are given by a probability function (rather than a probability measure). Although this is an important foundational issue, it will not be of central importance until we consider infinite probability spaces.

The Adams hypothesis is sometimes presented in a slightly different manner. Mellor (ibid., p. 234), for example, presents the Adams hypothesis for acceptance as follows: the acceptance of ‘If a, then b’ is given by \( cp(B|A) \approx 1 \), where ‘\( \approx \)’ indicates that the value of the conditional credence is at least close to 1, when \( cp(B|A) \) is defined.\(^{28}\)

How close the conditional credence \( cp(B|A) \) is (or should be) to 1 is not an issue that matters in these formulations. It would be useful briefly to compare this view with the one just sketched above. First, the improved Adams hypothesis for acceptance does not appeal to \( cp(B|A) \), but to a primitively defined notion of conditional probability. An agent can have definite views about the acceptance conditions of ‘If a, then b’, even when \( P(A) \) is zero.\(^{29}\) The original thesis cannot deal with those cases, while the improved version can. Even when the conditional credence \( cp(B|A) \) is undefined, the agent can have a view as to the degree to which \( B \) is expected upon supposing \( A \). And this can be determined by \( P(B|A) \). Second, the alternative formulation is equivalent to a view according to which full acceptance of ‘If a, then b’ requires that the conditional probability of \( B \) given \( A \) is infinitesimally close to one. In other words the improved acceptance test can be alternatively formulated by appealing to infinitesimal probability:

\(^{28}\)Mellor presents this version of Adams for expository and argumentative purposes. The positive view he defends differs from the one articulated in this version of the Adams hypothesis, and to some extent from the one presented here, although there are various important coincidences between Mellor’s view and mine, which will be discussed below.

\(^{29}\)Mellor makes an interesting point about acceptance (op. cit., p. 236). The use of a primitively defined notion of conditional probability is usually invoked in order to deal with the acceptance conditions of conditionals whose antecedents have measure zero. But there are other important set of cases, namely, all future conditionals used in decision making. The evaluation by Oswald of ‘If Oswald doesn’t kill Kennedy, someone else will’ can be an important part of Oswald’s process of deciding whether to make true the proposition \( O \) that he does not kill Kennedy. For \( O \)’s truth “depends on his decision, which he has not yet made and which we may consistently suppose him quite unable to predict. That is, he has and believes he has no credence for \([O]\) high or low or even indeterminate” (op. cit., p. 236). Yet Oswald can accept of ‘If Oswald doesn’t kill Kennedy, someone else will’, and this mental act might have an impact in his process of decision making. It seems that this type of conditionals, ‘If \( h, q \)’, where the agent has no view as to the value of \( P(H) \), cannot be straightforwardly handled by an unmodified version of unified probabilism. Possible solutions (within the boundaries of unified probabilism) will be discussed below.
Nonstandard probability test for acceptance: the simple conditional ‘If $a$, then $b$’ is accepted with respect to a nonstandard value probability function $p$ if and only if $\text{std}(p(B|A)) = 1$, where $\text{std}(x)$ is the unique standard number which differs from $x$ by at most an infinitesimal amount.\textsuperscript{30}

Third, the adoption of a primitively given notion $P(\_\_\_\_\_\_\_\_)$, or an infinitesimal probability function $p$, eliminates the need for the ‘otherwise’ clause in the formulation of the Adams hypothesis. This has various beneficial consequences, which will be the object of attention in the following paragraphs. There are two standards of validity in probabilistic semantics. One of them is genuinely probabilistic in kind.\textsuperscript{31}

Logic of likelihood criterion: an inference is probabilistically valid if and only if, for every positive $\varepsilon$, there exists a positive $\delta$ such that, under any probability assignment under which each of the premises will have probability greater than $1 - \delta$, the conclusion will have probability at least $1 - \varepsilon$.

There is a salient limit case by taking the limit when the uncertainties go to 0.

Logic of certainty criterion: an inference is strictly valid if and only if its conclusion has probability 1 under any probability assignment under which its premises each has probability 1.

The logic of likelihood criterion describes a plausible set of inferences excluding many suspect patterns validated by the material conditional (for example, the so-called “paradoxes of material implication”). Nevertheless, if one appeals to the original Adams hypoth-

\textsuperscript{30} In “Learning the Impossible,” McGee offered a back-and-forth theorem relating nonstandard functions $p$ and two-place functions $P$. Thomason and I extended this result to iteration in “Iterated Probability Kinematics,” Journal of Philosophical Logic (forthcoming). In the presence of both results, it is clear that for each $p$ used in the test of a conditional via the nonstandard probability test for acceptance, there is a two-place $P$ yielding identical acceptance conditions via the improved Adams hypothesis for acceptance. And, obviously, for each $P$ satisfying the improved Adams hypothesis for ‘If $a$, $b$,’ there is $\text{NS}(P)$ such that $P(B|A) = 1$ if and only if $\text{std}([\text{NS}(P)](B|A)) = 1$.

\textsuperscript{31} My terminology is the one used by McGee in “Learning the Impossible.”
esis, the strictly valid inferences are not the probabilistically valid inferences. The strictly valid inferences collapse into the materially valid inferences. As McGee points out: “in determining that the strictly valid inferences are the classical ones, what is important is [...] the default condition that assigns the conditional probability 1 when the conditional probability is undefined.”  

Then he shows that appealing to the improved hypothesis eliminates this anomaly. In fact, in this case the inferences probabilistically validated are exactly the ones that are strictly validated. An important conclusion follows. Consider the following pattern of inference $I$:

$$a^1, a^2, ..., a^n, a_i > b_j, ..., a_n > b_n \mid a > b$$

where $a^i, a_j, b_i, a,$ and $b$ are conditional-free formulae of the propositional language $L$. The letters $A, B, C, ..., \text{ stand for the propositions}$ expressed by the sentences $a, b, c, ..., \text{ of } L$. I shall also use an extension $LC$ of $L$ containing conditionals of the shape $a > \beta$ where $a$ is in $L$ and $\beta$ is in $LC$. $\alpha, \beta, \gamma, ..., \text{ stand for sentences of } LC$. Notice also that for sentences of $LC \setminus L$ there is no defined corresponding propositions. This reflects agnosticism about the existence of conditional propositions.

McGee has shown that any inference exhibiting the shape displayed in $I$ is probabilistically valid if and only if it is strictly valid, as long as we use the improved hypothesis. This is a powerful reason to argue that the improved Adams hypothesis is the most parsimonious rendering of the original Adams hypothesis. I shall argue that there are additional advantages associated with this formulation. In fact, I shall argue that the alignment between strict and probabilistic validity permits us to represent valid probabilistic inferences in terms of purely qualitative operations on the beliefs naturally associated with each two-place function. This representation makes it possible to bridge an important gap separating probabilistic and traditional epistemology.

My argument has two steps. The first step is simple. Strict validity, and, therefore, probabilistic validity, can be cashed out in terms of our test of acceptance. In fact, in order to check strict validity of a pattern $I$, one need only verify that the conditional $a > b$ is accepted

---

33 Given an underlying space $S = \langle U, F \rangle$ with $U$ a nonempty set and $F$ a sigma field on $U$, propositions are constructed as subsets of $2^U$.
34 In other words, the function that takes lower-case letters and yields corresponding upper-case letters is a map taking sentences of $L$ into $2^U$, in such a way that for every sentence $a$ of $L$, $A$ is the proposition expressed by $a$.
35 See theorem 3, page 190, of “Learning the Impossible.”
whenever each item in the antecedent of \( I \) is accepted. And acceptance of any conditional \( a > b \) goes by the generalized conditional probability of \( B_i \) given \( A_i \) being 1.

The second step is to show that the acceptance of each conditional \( a > b \) with respect to a two-place probability function \( P \) is expressible in terms of purely qualitative operations on the bodies of belief associated with \( P \) and \( P[A] \): the result of updating \( P \) with \( A \). Therefore, I shall show that probabilistic validity can be expressed in terms of probability-based belief change. This is the hard part of the argument. Establishing this second part requires a prior understanding of the (problematic) connections between probability and belief as well as a previous grasp of the relationships between probability kinematics and belief change. The following section focuses on this issue.

IV. TWO-PLACE FUNCTIONS, FULL BELIEF, AND PROBABILISTIC CORES

A space is a pair \( S = \langle U, F \rangle \) with \( U \) a nonempty set and \( F \) a sigma field on \( U \). A one-place probability measure on \( S \) is characterized by Kolmogoroff’s axioms:

1. \( 0 \leq P(A) \leq 1 = P(U) \),
2. \( P(A \cup B) - P(A \cap B) = P(A) + P(B) \) (finite additivity),
3. If \( E_n: n = 1, 2, \ldots \), are disjoint, with union \( E \), then \( P(E) = \sum_n P(E_n) \) (countable additivity)

Conditional probabilities can be defined in the usual manner. Alternatively, one can take two-place probability functions as primitives. This can be done as follows: a two-place probability function \( P(\cdot \mid \cdot) \) on space \( S \) is a map \( F \times F \) into real numbers such that:

I. Reduction axiom: the function \( P(\cdot \mid A) \) is either a probability function on \( S \) or else has a constant value 1. The latter function will be called abnormal.

II. Multiplication axiom: \( P(B \cap C \mid A) = P(B \mid A) \cdot P(C \mid B \cap A) \), for all \( A, B, C \) in \( F \). \(^{36}\)

If \( P(\cdot \mid A) \) is a one-place probability function, \( A \) will be called normal and abnormal otherwise. Let ‘\(+\)’ mark exclusive disjunction of propositions: \( A + B = (A - B) \cup (B - A) \). When two-place functions are taken as primitives, one-place probability functions are defined as follows: \( pr(A) = P(A \mid U) \). Two facts are worth keeping in mind: if \( A \) is normal, so are its supersets, and if \( A \) is abnormal, so are

\(^{36}\) The axiomatic use of the multiplication rule has a long pedigree in the theory of probability. Harold Jeffreys, for example, assumes it as a fundamental axiom of probability (under the name ‘W. E. Johnson’s product rule’) in his Theory of Probability (New York, Oxford, 1961; third edition), p. 25.
its subsets. In van Fraassen's construction, the notion of the a priori is the opposite of the idea of abnormality. In fact, A is a priori for P if and only if \( P(A|X) = 1 \) for all X, if and only if \( U - A \) is abnormal for P. Now we can introduce the notion of superiority: A is superior to B (denoted \( A >_P B \)) if and only if \( P(A|A + B) = 1 \). Armed with this notion, we can now introduce the following crucial concept: \( K \) is a belief core (for \( P \)) if and only if \( K \) satisfies the following axioms:

\[
\begin{align*}
(A1) & \quad K \text{ is normal} \\
(A2) & \quad \text{if } A \text{ is a nonempty subset of } K \text{ while } B \text{ and } K \text{ are disjoint, then } A >_P B \\
(A3) & \quad \text{the complement of } K, U - K, \text{ is normal}
\end{align*}
\]

The following property follows from the axioms and definitions: (finesse) all nonempty subsets of \( K \) are normal. In addition, van Fraassen has shown that belief cores are nested. The proof makes essential use of (A2) and finesse. van Fraassen has also suggested the following epistemological interpretation. Say that an agent’s epistemic state is represented by a probability function \( P \). Say that \( P \) induces a system of cores with nonempty intersection \( B \). The proposal is to see \( B \) as the strongest proposition (fully) believed by the agent. For the moment, nevertheless, we cannot define the full beliefs of an agent as the propositions entailed by the innermost core of his system of cores (if the agent has cores). In fact, even if the agent has cores, such an intersection is not obviously guaranteed to exist. Nevertheless, the following lemma shows that a non-coreless function always has a nonempty intersection of cores.

*Descending chains:* the chain of belief cores induced by a non-coreless two-place function \( P \) cannot contain an infinitely descending chain of cores.\(^{38}\)

Therefore, with the help of this lemma we can use the following crisp definition of full belief (fitting van Fraassen’s epistemological interpretation of core systems): A is a full belief (for \( P \)) if and only if A is entailed by \( P \)'s innermost core (or A is a priori). The following section will be devoted to van Fraassen’s account and to slight modifications and extensions of it.

---


\(^{38}\) See my “Qualitative and Probabilistic Models of Full Belief,” in S. R. Buss, P. Hájek, and P. Pudlák, eds., *Logic Colloquium ’98: Lecture Notes in Logic XIII* (Natick, MA: Peters and ASL, 2000), pp. 25-43. The lemma of descending chains (lemma 2) is stated and proved in pp. 36-37, together with relevant historical remarks, pointing to related work in the field.
IV.1. Zero probability and epistemic possibility. van Fraassen’s construction has an interesting behavior in spaces whose universe is countable. Let a point \( p \) in the universe \( U \) of a space \( S \) be called heavy for a function \( P \) if \( pr(p) = P(p | U) > 0 \).

Lemma of heavy points: if a space \( S = (U,F) \) is such that \( U \) is countable and normal for a non-coreless function \( P \) defined over \( S \), the set of heavy points of \( P \) constitute \( P \)'s smallest core.\(^{39}\)

van Fraassen’s construction intends to avoid the identification of full belief and probability 1 in various useful models. Moreover, as McGee clearly explains, one of the main motivations behind the use of two-place functions is to avoid identifying epistemic impossibility and measure zero:

If a proposition is utterly impossible epistemically, we shall assign it the probability 0, but it is by no means the case that any proposition to which we assign the probability 0 will be epistemically impossible.\(^{40}\)

Nevertheless, as long as ‘epistemically impossible’ means “incompatible with the full beliefs for \( P \),” the characterization of full belief offered in the previous section yields an identification of probability zero and epistemic impossibility. At least this is so if one focuses on countable probability spaces—via the lemma of heavy points. Here is a concrete example giving intuitive support to the idea that the aforementioned identification is problematic (when the universe is countable).

Example: the sample space has a countable set of atoms, resulting from the following: in independent trials, a fair coin is flipped until we get a head, and then the trials stop. The set of possible outcomes is indexed by the number of tails \( X = (0, 1, 2, ..., n, ..., \Omega) \), where \( \Omega \) designates never stopping, that is flipping forever and seeing only tails. Evidently \( P(X \text{ is finite}|U) = pr(X \text{ is finite}) = 1 \) and \( pr(X = n) = 2^{(n+1)} \ (n = 0, 1,...) \), so that (obviously) \( pr(X = \Omega) = 0 \), and this is the only null event, apart from the impossible event.

Notice that, according to van Fraassen’s model, there is only one core for \( P: U - \Omega = C \), constituted by \( P \)'s set of heavy points. But it seems that \( P \) should have two “cores”: \( C \) and \( U \), and it also seems that the agent’s full beliefs should be represented by \( U \) rather than by \( C \).

\(^{39}\) The result is presented by Rohit Parikh and myself as lemma 3.1 in “Two-place Probabilities, Beliefs, and Belief Revision,” in Paul Dekker, ed., Proceedings of the Twelfth Amsterdam Colloquium (Amsterdam: ILLC, 1999), pp. 1-6.

\(^{40}\) “Learning the Impossible,” p. 179.
The agent might be “almost certain” that the outcome is finite, but not fully certain. The space could have admitted physically impossible events (say, that the coin after being tossed stops and stands still in mid-air). Any of those (abnormal) events would have carried zero probability, like $\Omega$. But it seems that there should be a distinction between physically impossible events and $\Omega$. It is clear that none of the former events should be deemed as seriously possible. It is less clear that the same criterion applies to $\Omega$. This intuition cannot be captured inside of an unmodified version of van Fraassen’s system for two reasons. First, according to condition (A3) of cores, the complement of a core should be normal. Therefore, $U$ can never be a core. Second, the definition of full belief mandates to interpret the innermost core as the set of full beliefs. In “Two-place Probabilities, Beliefs and Belief Revision” (op. cit.), Parikh and I show that both obstacles can be removed via a slight modification of the mathematical apparatus used by van Fraassen. Here is the reformulation we proposed: first, redefine a core as a subset $K$ of $U$ obeying:

* Normality: $K$ is normal

* Strong superiority condition: if $A$ is a nonempty subset of $K$ while $B$ is disjoint from $K$, $P(B|A\cup B) = 0$

These two conditions are enough to entail finesse. Moreover, it can be shown in general (op. cit., lemma 2.4) that there is always a smallest core (the intersection of cores) and a largest core (determined by the union of cores). In addition, since the requirement that the complement of a core should be normal has been abandoned, the union of cores can perfectly be the universe of the probabilistic space. Then the universe $U$ in the example presented above (accommodating all possible trials, including the “never stopping” event $\Omega$) is now a permissible core of the function $pr$ used in the example.

A second modification can be suggested by noticing that the lemma of heavy points guarantees that the innermost core of a function $P$ is the strongest proposition carrying probability 1 (as long as the space over which $P$ is constructed has a countable and normal universe $U$, for $P$). The intuition resulting from the coin-tossing example is that defining full belief as what is entailed by the innermost core of a two-place function $P$ conflicts with the aforementioned formal fact. The agent represented in the example might at most expect the outcome of a trial to be finite. But it seems excessive to require that she ought to believe fully that $X$ is finite. Using a terminology that has some currency in Bayesian epistemology, one can say that the agent can be at most “almost certain” that the outcome will be finite, but not fully certain. One possible solution is
to adopt the following definitions: \( A \) is a full belief (for \( P \)) if and only if \( A \) is entailed by \( P \)'s outermost core (or \( A \) is a priori). \( A \) is expected for \( P \) if and only if \( A \) is entailed by \( P \)'s innermost core.

To be sure, expectations, as defined above, are identical with probability 1 in countable spaces. Moreover, the notion of full belief collapses with the a priori in such spaces. So, one can perfectly well analyze the countable case armed with only two epistemological notions: probability 1 and the a priori. The innermost core encodes the strongest probability 1 proposition (via the lemma of heavy points). What seems unintuitive in the case under consideration is to maintain simultaneously the characterization of full belief as what is entailed by the innermost core. So we have abandoned such characterization in general. The following diagram can help us to present in a succinct manner the framework we are suggesting in the countable case:

![Diagram](image)

Focus first on the (shaded) leftmost rectangle, which represents the universe \( U \) of a countable space \( S \). The boldfaced ellipses represent the system of cores for a two-place function \( Q \) over \( S \). The largest core \( F \) encodes \( Q \)'s associated full beliefs. Now the subsystem of cores embedded in this largest core can be seen as a series of inductive

41 It is important to keep firmly in mind throughout this essay that all the introduced epistemic notions are part of the conceptual apparatus of the particular brand of radical probabilism under consideration. The notion of expectation might play a different role in the context of a pluralist epistemology.

42 One of the main differences between van Fraassen's model and the model proposed here is that van Fraassen does not focus on countable models. On the contrary, the countable case is of negligible interest for him. His motivation is to capture the doxastic commitments of physical theories, and infinite spaces are crucial for this application. The focus here is on applying unified probabilism to improve existing models of conditionals, and for this application the countable case is crucial. I shall argue in sections VIII and IX that Adams's ideas are hard to extend to infinite spaces.
expansions of Q's full beliefs ordered by strength. The innermost core I can then be interpreted as the strongest possible expansion of F. I is constituted by all of Q's heavy points, while all points outside I are "light" points of probability zero. But Q carries enough qualitative information to distinguish among those "light" points. First of all, there are probability zero events which are epistemically possible (those overlapping F), and there are probability zero events which are epistemically impossible (those not overlapping F). Of course, all fully believed events carry probability 1, but not every probability 1 event is fully believed (for example, any subcore of F carries probability 1, but is not fully believed).

It is not difficult to see that every core of a function Q carries probability 1, and, therefore, so does the largest core of Q—given that the space is countable. But the central feature of the notion encoded via the largest core is that it is robust with respect to suppositions, in the sense that it carries probability 1 and continues to carry this probability under any supposition. In fact, it is not difficult to see that \( p_r(F) = 1 \), and \( Q(F|X) = 1 \), for any X. This robustness with respect to suppositions is the idea captured by van Fraassen's technical term 'a priori'. The weaker doxastic notion entailed by the innermost core also carries probability 1, but it is not robust with respect to suppositions. For example, \( Q(I|A) = 0 \), where I is the innermost core in figure one and A is the rightmost cell in the partition of the leftmost diagram. I propose to call the notion encoded by the innermost core expectation (almost certainty).  

A final remark is needed before concluding this section. Notice that according to my account, only certain events of probability zero are entertainable as candidates for possible updates. Those entertainable events should be epistemically possible, in the sense that they should overlap the largest core F. Epistemic models in terms of a two-place probability function P give us no guidance for modeling revisions of the full beliefs for P. In order to do that, we would need a gradation of the states in the complement of the largest core F. This has been pictorially represented in the leftmost rectangle of figure one via several dotted ellipses. The structure and epistemological function of such gradation is the target of the so-called theories of belief revision. The reader unfamiliar with this literature can think about this ordering of the points in the complement of F as a new

43 Let F and A be two compatible events in U. Then \( F \cap A \) is the expansion of F with A. Each subcore C of F can then be seen as an expansion of F with C.

44 This terminological decision is grounded on an established notational practice in philosophical logic and computer science.
“system of cores” now centered in the outermost core \( F \). This grading is usually interpreted as manner of encoding the degrees of epistemic entrenchment of full beliefs. Such grading is quite different in nature from the one I am considering here. In fact, the cores for a function \( P \) are not viewed as an entrenchment relation for full belief, but as a set of inductive expansions of the union of cores (that is, the agent’s full beliefs) ordered by strength. The strongest expansion is viewed as the strongest set of maximally likely propositions or expectations. On the other hand, the propositions outside of the body of full beliefs are not entertainable. They are not candidates for updating. The system of cores for \( Q \) can, nevertheless, be interpreted in our model as yielding the epistemic entrenchment of the expectations in the innermost \( Q \)-core (if such core exists). If we see things in this way, then the nature of our modeling reveals an important feature of two-place functions, namely, the manner in which the entrenchment of expectations for a function \( Q \) is tightly connected to a set of possible expansions of the set of full beliefs for \( Q \) (the cores for \( Q \)). If we focus on the case where countable additivity is enforced, the set of “closest” \( A \)-points to the expectations of a function \( Q \) always exists as long as \( A \) is epistemically possible—even when \( A \) and \( I \) are incompatible. This set can be seen as the revision of the set of expectations for \( Q \) with \( A \), or, alternatively as the boldest inductive expansion of the body of full beliefs for \( Q[A](X|Y) \), where \( Q[A] \) stands for the revision of \( Q \) with the proposition \( A \). In neither case does this operation have anything to do with a “genuine” revision of the full beliefs for \( P \). This matter, rather than the one I analyze here, is the main topic of the so-called theories of belief change.

\section*{IV.2. On the dynamics of cores.}

The general questions under investigation in this section are the following: (1) Which is the structure of the system of cores \( S[A] \) of \( P[A](X|Y) = P(X|Y \cap A) \)? (2) If \( S \) is the system of cores for \( P \), which is the relationship between \( S[A] \) and \( S \)? This is the last issue concerning cores that needs to be addressed before considering the probabilistic constraints on acceptance (of conditionals) induced by two-place functions.

Observation (1): if \( S \) is the system of cores of \( P(\_|\_|\_|\_\_) \) and \( A \) is epistemically possible for \( P \), then all the elements of: 
\[
S[A] = \{C \cap A : C \in S \text{ and } C \cap A \neq \emptyset\}
\]
are exactly all the elements of the system of belief cores for \( P[A] \).46

\footnote{A “genuine” revision would require an input incompatible with the current corpus of full belief.}

\footnote{For a proof, see my “Hypothetical Revision and Matter-of-fact Supposition,” \textit{Journal of Applied Non-Classical Logic} (forthcoming). A web version is available via the...}
A picture of core change is given by the rightmost and center rectangles in figure one. The shaded rectangle in this figure is partitioned by three events. Let the partition cells be called $P_1$, $P_2$, and $P_3$ (where $P_1$ is the leftmost partition cell). The center picture represents the result of updating by $P_1$, the rightmost picture the result of updating by $P_3$. The basic idea is that when a system of cores is updated by an epistemically possible event $E$ the web of inductive expansions intersecting $E$ is completely preserved.

One of the many interesting consequences of observation (1) is that it can be established in a general manner without using either countable additivity or the restriction of the space to the countable case. It is, therefore, a fairly robust property of two-place probability functions that the inductive structure they induce on full belief is preserved without further refinements every time they are updated.

Some limit cases and new notation are needed at this point. $(P[A])[B]$ will be abbreviated by $P[A, B]$. Whenever $A$ fails to overlap the largest core of $P$, or when $P$ is abnormal, we set $P[A]$ to the abnormal $P$ assigning probability 1 of every event. In addition, the innermost core of $P[A]$, denoted by $Ex(P)$, is, in this limit case, arbitrarily set to $\emptyset$. If $F(P)$ denotes $P$’s largest core, we also set $F(P)$ to $\emptyset$ by convention.

IV.3. Selecting the right dynamics. Section IV.2 started with the question: Which is the structure of the system of cores $S[A]$ of $P[A]$ $(X|Y) = P(X|Y \land A)$? But one might quarrel about the relevance of the question itself. In fact, there is certain freedom in defining $P[A]$. Given any $P$ and any proposition $A$ one can postulate (or explicitly construct) a selection function $\phi$ yielding the “closest” function $\phi(P, A)$ needed to change $P$ with $A$.47 Positing that $P[A](X|Y) = P(X|Y \land A)$ is tantamount to focusing on a specific manner of characterizing probability change. This particular definition can perhaps be defended by its naturalness as well as by established practice in the field.48 The main reason for choosing it here will be provided in the coming section. The target of section IV is to offer a representation of the improved Adams hypothesis for acceptance in terms of probability-based change. I shall show that, if $P[A](X|Y) = P(X|Y \land A)$ and

47 This idea has been considered by Peter Gärdenfors in Knowledge in Flux (Cambridge: MIT, 1988), chapter 5; and previously by Wolfgang Spohn in “The Representation of Popper Measures,” Topoi, v (1986): 69-74.

the underlying space is countable, then \( P(B|A) = 1 \) if and only if the expectations for \( P[A] \) entail \( B \).

This result offers a complete qualitative characterization of the acceptance of non-nested probability conditionals for probability spaces of countable size. It also clarifies the nature of the (traditional) epistemological notions tacitly involved in acceptance tests of probabilistic kind. This suggests the interests of considering extensions of the model in order to study iteration where the characterization of \( P[...|A] \) remains unchanged. Perhaps there are interesting applications of alternative models where a selection function is used in order to characterize \( P[...] \). Nevertheless, it is not clear that such models play a central role in representing qualitatively the Adams hypothesis.

V. SIMPLE CONDITIONALS

Before considering the iterated case, it would be beneficial to focus first on the structure of the improved Adams hypothesis for acceptance, introduced in section II.

**Improved Adams hypothesis for acceptance:** a simple conditional \( a > b \) is accepted with respect to \( P(-|--|A) \) if and only if \( P(B|A) = 1 \).

A similarly motivated idea can be expressed via an acceptance test closest in spirit to the so-called “Ramsey tests” used in recent epistemic models of conditionals.\(^49\)

\( \text{Ramsey test for simple probability conditionals:} \) a simple conditional \( (a > b) \) is accepted with respect to \( P(-|--|A) \) if and only if the smallest core of \( P[A] \) entails \( B \).

Now the act of supposing is qualitatively represented in terms of operations on propositions. In fact, the smallest core of \( P[A] \) yields the expectations for \( P[A] \). So, \( (a > b) \) is accepted if and only if \( B \) is hypothetically expected upon supposing \( A \).\(^50\) Notice that in this case

\(^{49}\) See the section devoted to epistemic models of conditionals in Corss and Nute; see also Ramsey’s “General Propositions and Causality,” p. 154, footnote.

\(^{50}\) In footnote 29 above, I mentioned an example (presented by Mellor) about the use of future conditionals used in decision making. An agent might accept one of those future conditionals ‘If \( a, b \)’ even when he has no prior \( P(A|U) \). Cases of this sort cannot be handled by an unmodified version of the improved version of the Adams hypothesis. In fact, using the system of cores induced by a (total) primitive two-place function \( P \) in order to deliver acceptance conditions for ‘If \( a, b \)’ would require having a defined value for \( P(A|U) \), even when this value is zero. So, Mellor’s challenge to the original Adams hypothesis seems to be also a challenge for the improved version of the hypothesis. A possible solution within the boundaries of probabilism would be to represent the consistent suppositions of the agent via a
the numerical properties of \( P \) only play a derived role. They supply the structure needed in order to generate a system of cores. Such system is the only structure required in order to implement the qualitative test. In other words, if we were given only the system of cores for \( P \), this would be sufficient to use the qualitative test. All other representational properties of \( P \), over and above its induced system of cores, are unnecessary in order to determine acceptance via the qualitative test. The qualitative test and the Ramsey test do not coincide in general. Nevertheless, they do coincide for countable probabilistic spaces.

**Coincidence lemma:** if the universe \( U \) of the underlying space is countable and normal for a function \( P \), then for all propositions \( A, B \), \( P(B|A) = 1 \) if and only if the smallest core of \( P[A] \) entails \( B \).

It is interesting to notice that the Ramsey test for simple conditionals is perfectly well defined for infinite spaces with more than countable points. In fact, the lemma of descending chains guarantees the existence of the smallest core for \( P \) and for \( P[A] \), as long as \( A \) overlaps the largest core of \( P \). Nevertheless, there is no guarantee that the Ramsey test and the improved Adams hypothesis should coincide for partial two-place function \( P \)—'consistent' here means that there is at least a core system delivering acceptance conditions for all the represented suppositions. Or we can use a set \( S \) of (total) two-place functions sharing the defined values of the partial function \( P \). If the intersection \( B \) of the propositions representing full beliefs for each total two-place function in \( S \) is nonempty, one can also infer that such proposition \( B \) represents the full beliefs for the partial two-place function \( P \).

51 Here is a sketch of the proof of the coincidence lemma. Assume \( Ex(P[A]) \subseteq B \). Then if \( F(P) \cap A \neq \emptyset \), we have that \( Ex(P[A]) \) is nonempty (by the lemma of descending chains) and \( P[A](Ex(P[A])) \cap U = 1 \). Therefore, the assumption yields \( P(Ex(P[A])) = 1 = P(A[B]) \). When \( F(P) \cap A = \emptyset \), \( P[A] \) is abnormal. Therefore, \( P[X|A] = 1 \) for all \( X \). In particular \( P[B|A] = 1 \). Assume now \( P[B|A] = 1 \). Then \( Ex(P) \) is the support \( S(P) = \{ x \in U : P(x) > 0 \} \)—by the lemma of heavy points. So, if \( A \cap F(P) \neq \emptyset \), we have \( Ex(P[A]) = S(P[A]) \)—by observation (1). Assume then by contradiction that \( Ex(P[A]) \) is not included in \( B \). The proof can then be completed by cases. First, consider the case \( Ex(P[A]) \cap B \) is empty. In this case, for all \( y \) in \( B \), \( P[A](\{y\}|U) = 0 \)—again by heavy points. Now, since the space is countable and normal for \( P \), countable additivity guarantees that \( 0 = \sum P[A](\{y\}|U) = P[A](B|U) = P(B|A) = 1 \). Contradiction. Second, consider the case when \( Ex(P[A]) \cap B \) is nonempty. In this case, there is a strict subset \( Z \) of \( Ex(P[A]) \), such that \( 1 = P[B|A] = P[Z|A] < 1 \). Contradiction. Third, in order to finish the proof, we should consider the case \( A \cap F(P) \) is empty. This case is immediate, because by definition \( Ex(P[A]) = \emptyset \subseteq B \).
spaces other than the ones we are studying here (countable spaces). I shall suggest (in section VIII) that the Ramsey test might be better able to preserve the intuitions of the Adams hypothesis for acceptance in those spaces. Concerning the countable case, the offered analysis supports the idea that simple probability conditionals should be understood as *expectation conditionals* of the shape ‘If a, then it is expected that b.’

Different variants of this idea have been defended in recent papers and monographs.

I warned the reader in the introduction that, when considered in general, the qualitative properties of belief change induced by two-place functions are at odds with some of the established standards in the field. This is more apparent once iteration is allowed. Nevertheless, a first discrepancy from the standard models appears at the noniterated level. In fact, these models are not *consistently preserving* in general. This property requires that the update of a consistent view with a consistent input should generate a consistent output. But when the model is not universal (that is, when the largest core does not coincide with $U$), updates with nonentertainable propositions yield the abnormal function. This nonclassical feature will be more evident by studying iterated models. This will be done in section VI.

**VI. ITERATION AND NEGATED CONDITIONALS**

My immediate concern here will be to make precise what a probabilistic model is. The first component of the model is a probabilistic space $S = (U,F)$. I shall consider tuples of the form $M = (S, C, (Ex, F), Sup)$, where $S$ is a space, $C$ is a set of non-coreless two-place probability

---

52 This is so, at least, for spaces whose universe is normal for the relevant function $P$ constructed over them. The important special case when the largest core coincides with the universe $U$ of the space is studied in “Hypothetical Revision and Matter-of-fact Supposition.” The resulting model for (one shot) change of the innermost expectations is, in this setting, well known. It obeys the so-called Alchourrón-Gärdenfors-Makinson axioms, standardly assumed in the field—see Gärdenfors. Parikh and myself have also shown (op. cit.) that the conditionals validated by the Ramsey test satisfy the axioms of the theory of nonmonotonic reasoning presented by D. Lehmann and M. Magidor in “Rational Logics and Their Models: A Study in Cumulative Logics,” Technical Report TR 88-16, Department of Computer Science, Hebrew University of Jerusalem (November 1988).

53 Gärdenfors and David Makinson, “Nonmonotonic Inference Based on Expectations,” *Artificial Intelligence*, lxv, 2 (January 28, 1994). A similar idea appears also in several papers by Levi, as well as in his *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning* (New York: Cambridge, 1996). See also Andre Fuhrmann and Levi, “Undercutting and the Ramsey Test for Conditionals,” *Synthese*, cl, 2 (1994): 157-69. Nevertheless, the similarities between these proposals and the one suggested above should not be overstressed. Detailed comparisons (which are beyond the scope of this article) can be found in “Two-place Probabilities, Beliefs, and Belief Revision.”
functions defined over $S$ obeying the axioms I and II, and $\text{Sup}$ is a function from $C$ to $T$, the theories definable over $LC$. The pair $(\text{Ex}, F)$ contains the function $\text{Ex}$, which, when applied to any $P$ in $C$, yields the smallest core of $P$ (intuitively, $P$’s expectations) and the function $F$, which, when applied to any $P$ in $C$, yields the largest core of $P$ (intuitively, $P$’s full beliefs). Finally, [...] is a function mapping probability functions on probability functions. We do not require the closure of $C$ under [...] . A notational remark: so far I have used the letters $A, B, C,...$, in order to denote the propositions expressed by the letters $a, b, c,...$, of $L$. In addition, let $LC$ be the smallest language such that $L \subseteq LC$ and if $a \in L, \beta \in LC$, then $a > \beta \in LC$ and $\neg \beta \in LC$. From now on we shall assume that $A_M$ is the proposition expressed by $a$ in model $M$. I shall also use the constants true and false, and assume that, for every model $M$, $\text{true}_M = U$.

Various restrictions on those models have been proposed. First, there are the countable models, where the universe $U$ of the space $S$ contains countable many elements. Second, there are the universal models where, for each $P$ in $C$, the largest core for $P$ is $U$. Third, there are the normal models, where all points in $U$ are normal for any $P$ in $C$. In addition, all models obey:

**Ramsey test for probability conditionals:** $a > \beta \in \text{Sup}(P)$ if and only if (1) $b = \beta$ is in $L$ and the smallest core for $P[A]$ entails $B$, or (2) $\beta \in \text{LC} \setminus L$ and $\beta \in \text{Sup}(P[A](X|Y)) = \text{Sup}(P(X|Y \cap A))$.

**Ramsey Test for negated probability conditionals:** $(a > \beta) \in \text{Sup}(P)$ if and only if (1) $b = \beta$ is in $L$ and the smallest core for $P[A]$ does not entail $B$, or (2) $\beta$ is in $\text{LC}$ and $\beta \not\in \text{Sup}(P[A](X|Y)) = \text{Sup}(P(X|Y \cap A))$.

The basic idea of the test for negated conditionals is that the acceptance of $\neg(a > \beta)$ reveals that the agent thinks that $\beta$ is not accepted in the minimal hypothetical change needed to suppose consistently that $A$ is the case. In probabilistic terms when $\beta = b$, the idea is that $\neg(a > \beta)$ is accepted with respect to $P$ whenever $P(B|A) \neq 1$. For decisive probability functions our proposal is tantamount to establishing that $\neg(a > b)$ is accepted with respect to $P$ whenever $P(B|A) = 0$. But, in my framework, this is just a limit case.\(^{55}\)

---

\(^{54}\) And where the derived one-place functions obey axioms 1, 2, and 3.

\(^{55}\) Notice that the model precludes being in suspense toward conditionals. This epistemic feature can be relaxed by ceasing to require the test for negated conditionals.
A conditional $a > \beta$ is strictly valid in a model $M = \langle S, C, \langle Ex, F \rangle, Sup \rangle$, if and only if $a > \beta \in Sup(P)$ for all probability functions $P$ in $C$. A conditional $a > \beta$ is strictly valid if it is valid in all models. Finally, a conditional $a > \beta$ is strictly valid in a class of models $E$ if it is valid in all $E$-models. Strict validity can be extended to arbitrary formulas of $LC$: a sentence $\alpha \in LC$ is strictly valid in a model $M = \langle S, C, \langle Ex, F \rangle, Sup \rangle$ if and only if $\alpha \in Sup(P)$ for all probability functions $P$ in $C$.\textsuperscript{56}

Intuitively, $Sup(P)$ yields the set of conditionals accepted (or supported) by each probability function $P$ in $C$ together with all the logical consequences of such set. The interpretation of nonconditional sentences of $L$, contained in $Sup(P)$, is an entirely different matter. Ultimately, the propositions expressed by these sentences are propositions accepted by the represented agent. Various forms of radical probabilism will, of course, represent acceptance of propositions in terms of high probability. In the present setting, this will require stipulating that proposition $X$ is accepted with respect to $P$ if and only if $X$ is entailed by the smallest core of $P$ (or, in other terms, if $X$ is expected). It is not surprising perhaps that this is exactly Adams’s position in this matter, given that this view is compatible with the original Adams hypothesis. In fact, Adams has recently suggested that we treat equally the information carried by $A$ and by $true > a$. In other words, he suggested requiring $a \leftrightarrow true > a$.\textsuperscript{57} This can be captured in our theory by imposing the following additional postulates: (1) If $a \in L$ and $a \in Sup(P)$, then $Ex(P) \subseteq A$, (2) $Th(Ex(P)) \subseteq Sup(P)$—where ‘$Th(Ex(P))$’ yields the $L$-theory corresponding to the proposition $Ex(P)$. These two additional constraints essentially tell us that all nonconditional information supported by a two-place function $P$ is constituted by the expectations for $P$.

\textsuperscript{56} Evaluating the probability of mixed formulae like $a \rightarrow (true > a)$ is a complicated matter, both in the case in which conditional propositions are admitted and when they are not. The present proposal circumvents the problem by appealing to the syntactic closure of support sets. So, checking the strict validity of $a \rightarrow (true > a)$ requires verifying whether the formula belongs to all possible support sets. It is noteworthy, however, that even when the offered model provides acceptance conditions for mixed formulas and embedded conditionals, it does not assume that these formulas have probabilities. In this respect it resembles acceptance models for conditionals in terms of high probability—see, for example, Adams’s model in “On the Logic of High Probability,” Journal of Philosophical Logic, xv (1986): 25579. So, for example, if $\alpha$ and $\beta$ are simple conditionals, $(\alpha \vee \beta) \notin Sup(P)$, indicated that neither conditional is accepted with infinitesimally high probability, and where the sentence $(\alpha \vee \beta)$ is not assumed to carry probabilities. In addition, the offered model articulates the doxastic commitments involved in acceptance tests in terms of high probability. Acceptance tests for simple conditionals can characterize degrees of acceptance. More complex models can only deploy the attitudes of acceptance and rejection.

\textsuperscript{57} See A Primer of Probability Logic, p. 154.
It is clear that \((true > a) \rightarrow a\) is strictly validated in the presence of constraint two—it is assumed here that \(a \in L\). Nevertheless, the converse is not strictly valid. Any function \(P\) whose associated expectations contain both \(A\) and \(\neg A\) points constitutes a counterexample. Nevertheless, \(a \rightarrow (true > a)\) is negatively validated in our models as long as constraint one is imposed. A sentence \(\alpha \in LC\) is negatively valid in a model \(M = (S, C, (Ex, F), Sup)\) if and only if \(\neg \alpha \notin Sup(P)\) for any probability function \(P\) in \(C\). That the notions of negative and strict validity come apart in probabilistic models of the type we are considering should be clear from the status of \(a \rightarrow (true > a)\).\(^58\)

Now, of course, characterizing acceptance of propositions in terms of “almost certainty” is not an appealing option for unified probabilism. In fact, the view of acceptance in terms of “almost certainty” is exactly the naïve view of acceptance that this new form of radical probabilism tries to reform, given that it is affected by lottery paradoxes. It is more reasonable to characterize acceptance in terms of full belief, that is, in terms of what is entailed by the largest core of a two-place probability function. Adopting this point of view is compatible with the endorsement of the following two constraints, instead of constraints (1-2): (1f) if \(a \in L\) and \(a \in Sup(P)\), then \(F(P) \subseteq A\), (2f) \(Th(F(P)) \subseteq Sup(P)\).

If we adopt this second option, the relationships between \((true > a)\) and \(a\) become more tenuous. Only \(a \rightarrow (true > a)\) remains negatively valid. But the converse does not hold, due to the fact that expected items need not be fully believed. It seems that this second option is the one compatible with the endorsement of the “improved Adams” and with the tenets of unified probabilism.

VI.1. Probabilistic and epistemic models. The model presented in section vi applies to embedded conditionals and to negations of probability conditionals. Many probabilists have defended the validity of inferential patterns of this sort, even when their arguments tend to appeal to possible-worlds models or to other nonprobabilistic machinery. The present model also articulates probabilistic acceptance in terms of purely qualitative operations on propositions; but those operations are probabilistically justified via the connection between qualitative and quantitative doxastic states presented in section iv.

The reader can still ask: What is the place of the Ramsey test for probability conditionals in the space of current theories of condition-

als? The test is a garden-variety of “Ramsey test” recently used in the
theory of epistemic conditionals. But it is a very special type of
Ramsey test. By the same token, one can ask what is the relationship
between the test and the orthodox versions of probability logic. Some
of these questions have already been considered above, at least
partially, but I now want to review the complete answers.

All variants of Ramsey tests determine acceptability with respect to
an epistemic state E, which can (but need not) be a proposition. $a > b$
is accepted with respect to E if $B$ is accepted in the hypothetical
transformation of $E$ with $A$. Different versions of the test are deter-
dined by different choices of the underlying epistemology, which, in
turn, determines the nature of $E$ as well as the nature of the suppo-
sitional operation on $E$. Therefore, different types of conditionals can
be parametrically classified by varying the background epistemology
used in the formulation of their acceptance conditions. The test
presented above is importantly constrained by an underlying episte-
omology of probabilistic kind, where the only primitive is a notion of
generalized conditional probability. Once this epistemology is
adopted, the improved Adams hypothesis for acceptance can be
formulated in purely qualitative terms, but this formulation induces a
strong constraint on the nature of the epistemic state with respect to
which acceptance is judged. The intended targets are the expectations
for the primitive two-place function $P$, rather than the full beliefs for
that function. The expectations for $P$ are also the body of qualitative
attitudes that need to be hypothetically transformed in order to
evaluate probabilistic conditionals. On the other hand, the underly-
ing probabilistic epistemology imposes important constraints on this
notion of supposition, which other nonprobabilistic operations on
primitively defined expectations do not need to assume. In section
VII, I shall axiomatize the relevant notion of hypothetical change and
compare it with other characterizations of supposition in the litera-
ture. I shall also show that this notion of change is inconsistent with
many of the well-known notions of supposition used in the literature.
The singularity of the Ramsey test for probability conditionals is in
part determined by the singularity of the notion of supposition
induced by the underlying probabilistic epistemology.

I now turn to probability logic. The cornerstone of the contempo-
rary theory of probability conditionals is the idea (the Adams hypoth-
esis) that probability of conditionals is conditional probability.
Perhaps the central lesson of Lewis’s “Probabilities of Conditionals
and Conditional Probabilities" (op. cit.) is to make clear that the Adams hypothesis is incompatible with the idea that conditionals express propositions. So, there is a clear choice between the Adams hypothesis and truth conditionality. Adams and his associates have chosen to give up truth conditionality in order to preserve the hypothesis. The choice is motivated by the plausible idea that the assertability of conditionals goes by conditional probability. This insight is a highly entrenched tenet of probabilistic semantics. Truth conditionality, on the other hand, is a thesis which probabilists are ready to abandon on the basis of recalcitrant evidence.60

Here we should face an important issue concerning my present proposal. As Adams has made rather clear in recent writings, the important problem in probabilistic semantics is not only how to capture the inference patterns validated by probabilistic conditionals. Many rival theories might be able to satisfy this requirement. The crucial problem is to produce a theory capable of capturing these patterns as well as the insight that assertability of simple (and contingent) conditionals goes by conditional probability. Adams made this point in A Primer of Probability Logic while commenting on Stalnaker's possible-worlds semantics.61

While the Stalnaker theory has the same consequences as the probabilistic theory so far as the validity of inferences is concerned, the probabilities assigned to conditionals by it are not conditional probabilities, and therefore it disagrees with the probability conditional theory at the level of probability (op. cit., p. 196).

Since Stalnaker assumes truth conditionality, the probability of his conditionals cannot go by conditional probability. And this is so even when there is a coincidence at the level of validity for inferences involving simple conditionals.

The notion of strict validity used here is also capable of capturing the inferential patterns sanctioned by the probabilistic theory. But unlike rival views, there is no disagreement with the probability conditional theory at the level of probability. The present proposal

60 The dominant view among contemporary probabilists is a moderate agnosticism about truth conditionality. Perhaps there is a coherent theory showing that conditionals are truth bearers, but such theory is not needed in order to propose assertability conditions for conditionals understood as syntactical objects. Other scholars from Ramsey to Allan Gibbard have been more emphatic in denying that conditionals express propositions.

follows the probabilistic theory in not assuming truth functionality. Therefore, it is perfectly compatible with the idea that the degrees of assertability of conditionals go by conditional probability—to the extent that this idea can be coherently formulated.

It is important to notice that, although the acceptability of \(a > b\) is determined by \(P(B|A)\) assuming the value 1, my theory does not appeal to decisive functions with values ranging exclusively in \([0, 1]\). McGee\(^{62}\) has recently used these functions in order to capture Stalnaker’s logic via a probabilistic model. Decisiveness is permitted but not obligatory in our models. The crucial difference between models like McGee’s and my present account is the appeal to a slightly modified version of van Fraassen’s theory. Corpora of expectations and full belief are determined for all functions in the model, including nondecisive ones.

VI.2. Ramsey test for full belief. The qualitative test for simple probability conditionals can be formulated for other attitudes rather than expectations:

\textit{Ramsey test for two-place functions}: a simple conditional \((a \Rightarrow b)\) is accepted with respect to \(P(\rightarrow -)\) if and only if the largest core of \(P[A]\) entails \(B\).

According to this test, the conditional \(a \Rightarrow b\) is accepted with respect to a two-place function \(P\) if and only if as many as possible of the full beliefs for \(P\) as are compatible with \(A\) entail \(B\). So, the Ramsey test for two-place functions is closer in spirit to the various acceptance tests recently considered in the literature devoted to epistemic conditionals (where tests for acceptance of conditionals are usually formulated in terms of suppositional operations on bodies of full belief). In addition, the conditional patterns induced by the Ramsey test for two-place functions diverge from the ones validated by the Adams hypothesis for acceptance. Here is an example to substantiate this claim. The conditional \((\text{True} > \neg a) \rightarrow (a > \text{False})\) is not strictly valid, while \((\text{True} \Rightarrow \neg a) \rightarrow (a \Rightarrow \text{False})\) is. Even when the acceptance of \((\text{True} > \neg a)\) determines that the unconditional probability of \(A\) is zero, only the acceptance of \((\text{True} \Rightarrow \neg a)\) is sufficient to determine the serious impossibility of \(A\), in the sense that \(P[A]\) is abnormal. Acceptance of \((\text{True} > \neg a)\) is perfectly compatible with the normality of \(P[A]\). This reinforces the claim made above to the extent that probability conditionals are, in an important sense, expectation con-

\(^{62}\) Actually, McGee’s result captures a non-nested fragment of Stalnaker’s logic.
To be sure, one can study the logic of the conditionals induced by the Ramsey test for two-place functions, but these conditionals have a different logic from the one studied in probability logic.

VII. HYPOTHETICAL REVISION

We are now in a position to see the main properties of qualitative supposition induced by two-place functions. Consider a model $M = \langle S, C, \langle Ex, F \rangle, Sup \rangle$, the following postulates are satisfied by non-coreless functions $P$.

- **Entailment:** $Ex(P) \subseteq F(P)$
- **Full belief expansion:** $F(P) \cap A = F(P[A])$
- **Success:** $Ex(P[A]) \subseteq A$
- **Preservation:** if $Ex(P) \cap A \neq \emptyset$, then $Ex(P) \cap A = Ex(P[A])$
- **Restricted consistency preservation:** if $A \cap F(P) \neq \emptyset$, then $Ex(P[A]) \neq \emptyset$
- **Entertainability:** if $F(P) \cap A = \emptyset$, then $P[A]$ is abnormal
- **Fixity:** if $P$ is the abnormal function, then $Ex(P[A]) = F(P[A]) = \emptyset$, and $P[A] = P$
- **Cumulativity:** $Ex(P[A, B]) = Ex(P[A \cap B])$

Some of these postulates follow from basic definitions, others are corollaries of previous results. For example, both the proof of cumulativity and the proof of preservation require substantial use of observation one, while restricted consistency preservation is a consequence of the lemma of descending chains. The notion characterized by the previous postulates will be called from now on hypothetical revision.

It is worth noticing that full belief expansion, together with fixity, entails the converse of entertainability, which we can call consistency. In other words, the nonentertainability, or serious impossibility, of a proposition $A$ for $P$ (that is, the fact that $F(P) \cap A = \emptyset$) is both a necessary and a sufficient condition for the abnormality of $P[A]$. While consistency is a widely accepted Bayesian principle, most qualitative theories of belief change allow for the formation of policies for changing one’s view upon learning (supposing) that $A$ when the agent is certain that $A$ is false. The probabilistic tradition based on the use of Popper functions is not equipped to deal with such full-belief-contravening changes.

\[63\] It is also important to realize that no preferential (nonmonotonic) logic of the type studied by Lehmann and Magidor (op. cit.) validates the pattern $(true > \neg a) / (a > false)$. This also reinforces the idea that most nonmonotonic notions of consequence can be represented as expectation conditionals where the term ‘expectation’ is technically understood in probabilistic terms.

\[64\] See my “Hypothetical Revision and Matter-of-fact Supposition,” as well as Thomason and myself.
So, one can arrive at the former characterization of probability-based belief change by defining \( P[A](X|Y) \) as \( P[A](X|Y \cap A) \), and by focusing on the properties of core systems. Alternatively, hypothetical revision can be adopted as a primitive, without specifying the meaning of \( P[A](X|Y) \) via an explicit definition. It is easy to see that there is more than one notion of probability change compatible with hypothetical revision. On the other hand, there are quite robust properties entailed by the axioms just proposed. For example, the following property is entailed just by “full belief expansion”:

Cumulativity for full belief: \( F(P[A, B]) = F(P[A \cap B]) \)

Cumulativity for full belief is an important qualitative property induced by the type of radical probabilism under consideration here. In fact, there is evidence that cumulativity is a central ingredient in the research program whose only primitives are either infinitesimal probability or two-place functions. Thomason and I (op. cit.) provide such evidence by showing that cumulativity for full belief is needed in order to extend to the iterated case the mapping between infinitesimal probability and extended conditional probability first developed by McGee.

All the theoretical pieces of the puzzle are now in place in order to substantiate one of the main claims made in the introduction above. One of the challenges presented there was to show that the so-called export-import law is a deeply entrenched commitment of the form of radical probabilism used in order to improve Adams’s original hypothesis.

**Export-import:** \( \phi > (\psi > \chi) \iff (\phi \land \psi) > \chi \), where ‘\( \iff \)’ is the standard biconditional and \( \phi, \psi \in L, \chi \in LC \).

The connections between cumulative models, exportation, and importation are articulated in the next section, together with other properties of probability conditionals induced by hypothetical revision.

**VII.1. Exportation and importation.** It is easy to see that the export-import law is strictly validated in the class of normal and countable models. This is guaranteed by the coincidence lemma established in section v (and proved in footnote 51), and by the definition \( P[A](X|Y) = P(X|Y \cap A) \).65

---

65 To see that notice that the acceptance of \( a > (b > c) \) with respect to \( P \) depends on whether \( C \) is accepted with respect to \( (P[A])[B] \). Since by definition \( (P[A])[B](X|Y) = P(X|Y \cap A \cap B) \), we have that \( a > (b > c) \) is accepted with respect to \( P \) as long as \( P(C|A \cap B) = 1 \).
The result can be generalized in an interesting manner. In fact, two assumptions used above can be considerably weakened. On the one hand, both exportation and importation hold in a larger class of uncountable models. The following section will focus on those models. On the other hand, the export-import law is valid in all cumulative models even when $P[A][X|Y]$ is not defined as $P(X|Y\cap A)$. This second point is particularly important given the centrality of cumulativity for the versions of radical probabilism under consideration. It is clear, as we saw above, that cumulativity is enforced if $P[A][X|Y]$ is defined as $P(X|Y\cap A)$. But there might be alternative definitions of $P[...]$ whose associated notion of probability-based change are cumulative. Both exportation and importation hold for all those definitions of $P[...]$.  \footnote{This can be proved directly. In fact, cumulativity guarantees that assuming the antecedent of each law yields the correspondent consequent. Notice that here we are assuming cumulativity as an axiom rather than relying on a particular definition of $P[...]$—and then using observation one.}

Although I am not focusing here on the conditionals induced by the Ramsey test for full belief, I can say in passing here that the export-import laws also hold for those conditionals. In fact, cumulativity for full belief is a rather robust feature of the semantic program based on the use of generalized conditional probability. Of course, $a > a$ (success) is immediately validated, and a straightforward consequence of having the export-import law and success is: (global success) $a > (b > a)$. It should also be mentioned that neither export-import nor global success is compatible with most of the standard theories of theory change. Intuitively, global success says that every input hypothetically accepted at a certain point is rigidly maintained in future changes. This is against the one-shot property of consistency preservation assumed by almost all theories of theory change. The idea of consistency preservation is that a consistent corpus changed by a consistent input should yield a consistent output. It is easy to see that the method induced by two-place functions is not consistency preserving (just consider $a > (\neg a > a)$). It is true, nevertheless, that theories of conditionals developed with an eye on the role of conditionals in hypothetical reasoning and decision making tend to assume variants of export-import without much hesitation (sometimes for mathematical convenience). For example, Richard Bradley\footnote{“A Representation Theorem for a Decision Theory with Conditionals,” Synthese, cxvi, 2 (1998): 187-229.} offered one of those theories in a recent article. The present model validates all of his axioms \textit{(ibid., p. 191)} with the
exception of the schema: \( \neg(a > b) \leftrightarrow (a > \neg b) \). This axiom is rather controversial, and Bradley accepts it only reluctantly. My model validates a weaker property in the presence of normality: (fullness) \( \neg(a > \gamma) \leftrightarrow (a > \neg \gamma) \), where \( \gamma \in LC \) and \( a \in L \).

VIII. INFINITE PROBABILITY SPACES

So far I have mainly focused on countable spaces. We have seen that certain conditional laws continue to be valid even if the countable restriction is lifted. Nevertheless, there are many reasons for being very cautious in incorporating uncountable spaces. Some of these reasons are technical and go deep into the foundations of probability theory. The gist of the problem is that according to the received view in conditional probability, if \( w \) is a point in \( U \), \( P(\neg | w) \) is not probability given an event, but probability given a sigma field. Some authors have formulated variants of the theory of two-place functions offered here, where relativizations are used. For example, Renyi’s characterization of conditional probability does appeal to a relativization, not to sigma fields but to bunches. Formally, \( B \) is a bunch of a space \( S \) if it is closed under finite unions, the empty set does not belong to \( B \), and there is a sequence of events in \( B \) whose union equals the universe \( U \) of \( S \). Intuitively, the bunch should group a set of observable events, and only those events should be admissible conditioning events. Unfortunately, Renyi does not offer a clear operational test for identifying observable events, although some researchers have recently proposed such tests in particular areas of application of the theory. We can easily adopt Renyi’s axiomatization here, but ultimately this does not seem sufficient to produce a smooth connection with contemporary measure theory. This is so, at least, when countable additivity is adopted and the space is uncountable. There is, nevertheless, no need to go deeper into this issue here. In fact, I shall suggest below that: (1) there seem to be good reasons for not focusing on uncountable models in the first place, and (2) the axiom of countable additivity is suspect for the application we are considering.

Although the gist of Adams’s hypothesis is conceptually appealing, there is no guarantee that it will continue to be conceptually compelling when applied to uncountable probability spaces. The notion of probability delivers a nice account of the notion of measure and size comparison, and it does so both in the infinite and finite case. But it is not clear that an epistemological interpretation attached to measures (as encoding degrees of belief or degrees of conditional belief) needs to be robustly maintained for spaces of different sizes. In fact, I would like to offer reasons against that view by defining an infinite conditional probability space, and proposing a concrete ex-
ample showing that the probability function so defined delivers un-
intuitive acceptance conditions.

Renyi provided several examples of conditional probability spaces
obeying both the reduction axiom and the multiplication axiom
introduced in section IV. For example, let $R$ be the real line, and let
$BR$ be the $\sigma$-algebra of Borel subsets of $R$. Let $\lambda(A)$ denote
the Lebesgue measure of a set $A \in BR$. For all $A$ and $B$ in $BR$ such that
$\lambda(B)$ is strictly greater than 0 and strictly less than positive infinite, we
can define $p(A|B)$ as $\lambda(A \cap B)/\lambda(B)$. We then get a full
probability space. This space does obey the reduction axiom and the multipli-
cation axiom. Consider now the following situation in which an agent
evaluates the weight of a piece of stone in the real interval $[0,1]$. We
can assume that the function $p$ encodes the agent’s epistemic state. Notice that, according to the above definition, the agent is commit-
ted to accept ‘If a number measures the weight in the $[0,1]$ interval,
then the number in question is irrational’. There are plenty of
examples of this kind—for example, ‘If the pointer is set up in such
a way that it might land randomly at any point in the real interval
$[0,1]$, then it will land at an irrational number’.

IX. COUNTABLE ADDITIVITY AND EPISTEMIC PARADOX

Up to this point, I have taken for granted that the underlying one-
place probability functions obey Kolmogoroff’s axioms. Is there any
reason for thinking that these classical axioms (stated in section IV)
might be problematic? Here I would like to argue that countable
additivity might be questionable.

One of the main goals of van Fraassen’s framework is to deal with
the so-called lottery paradoxx: say that agent $a$ assumes that the weight
of a stone is representable by a real number in some interval, say,
between $.5$ and $1$ pounds. $a$ is certain of the hypothesis that stipulates
that the weight lies somewhere in the interval. Yet $a$ might coherently
assign 0 probability to the uncountable many rival hypothesis of the
form, ‘The weight is exactly $x$ pounds, where $x$ is a real value between
$.5$ and $1$’. So, $a$ fully believes that each of these rival hypothesis will
not obtain, but he is also certain that one of them will obtain.

van Fraassen’s construction circumvents this type of situation by
denying that probability 1 is sufficient for full belief. Consider now
the following variant of the paradox of the lottery: take any effective
enumeration $r_1, r_2, ..., r_n, ...$ of the rationals in $.5, 1]$. Consider now
the following family $T$ of hypothesis: ($T1$) the weight is exactly $r_1$;
($T2$) the weight is exactly $r_2, ...$, ($T3$) the weight is exactly $r_n, ...$.

There are countably many of these hypotheses. It seems excessive
to dismiss as irrational the idea that the hypotheses in $T$ are equiprob-
able. Now, if the hypotheses in $T$ are equiprobable, they cannot carry positive probability (because finite additivity guarantees that some finite union of hypotheses in $T$ will carry more than unit probability). But one can assign 0 probability to each $T_i$. Then all the probability mass will concentrate in co-finite families of hypothesis in $T$. But now the assumption that probability 1 entails full belief leads to a trans-finite lottery paradox. For one should fully believe that each of the countably many rival hypothesis $T_n$ will not obtain, and one should also be practically certain that the weight lies somewhere in the rational interval $[.5, 1]$. Symmetry seems to require a uniform treatment of both countable and uncountable transfinite lottery paradoxes. van Fraassen’s theory does account for both paradoxes, but the account is asymmetric. The first paradox is resolved because in van Fraassen’s theory, probability 1 does not automatically entail full belief. The treatment of the second paradox is slightly more ad hoc. Since van Fraassen’s unconditional probability functions obey countable additivity, the scenario depicted in the former example is simply ruled out as incoherent. Of course, this is also a problem for the improved Adams hypothesis. In fact, consider: $(EQ)$ if the pointer is set up in such a way that it might land randomly at any point in the rational interval $[0,1]$, then the possible outcomes are equiprobable. It seems too extreme to rule out a priori the acceptability of $(EQ)$ as a form of epistemic incoherence. Nevertheless, this is exactly what the improved Adams hypothesis recommends, when the underlying unconditional probability is countably additive. To see that, notice that the proposition $E = \{x \in U : P(x|U) = 0\}$ and for every point $y \in U$ such that $y \neq x$, $P(|y|U) = 0$ is, according to the axioms, empty.

**IX.1. Finitely additive models.** A radical solution to the problem presented above is just to abandon countable additivity. In that case, the adoption of a uniform unconditional prior on a countable space is perfectly possible. And, of course, $EQ$ ceases to be analytically false with respect to finitely additive models.

Many crucial features of van Fraassen’s picture remain untouched when we adopt finite additivity. For example, each non-coreless function continues to induce a nested system of cores. Nevertheless, there is no guarantee of the existence of an innermost core. Therefore, the definition of expectations cannot be done in terms of the intersection of all cores. But this is not an obstacle against defining expectations. One can appeal to the type of definition used by van Fraassen himself. One can say that the proposition $A$ is expected (for
if and only if $A$ is either a priori or entailed by some core for $P$.68

So, for non-coreless $P$, we can define $\text{Exp}(P)$ as the set of all propositions expected for $P$.

$$\text{Exp}(P) = \{A : A \text{ is entailed by some core of } P\} = \text{Exp}(P)$$

This move still allows us to characterize the acceptance of simple conditionals qualitatively in terms of what is entailed by expectations. Basically, $a > b$ is accepted with respect to $P$ if and only if $B$ is in $\text{Exp}(P)$. And, of course, the definition of full belief remains unchanged. Nevertheless, an interesting symmetry is immediately broken. In the countable case, $P[A](B) = 1$ guarantees that some core for $P[A]$ entails $B$. Once countable additivity is abandoned, it is possible that there is no core for $P[A]$ entailing $B$ although $P[A](B) = 1$. As an example, assume that $P(B|A) = 1$ and set $B \subseteq A$. The following situation can then arise. One can have infinitely many cores $C$ (for $P$) cutting $B$ (but not included in $B$) in such a manner that the intersection of each $C$ with $B$ carries probability 1. So, the following test does not capture the qualitative counterpart of the Ramsey test for probability conditionals when countable additivity is abandoned:

$$\text{Expectation test for probability conditionals: } a > b \in \text{Sup}(P) \text{ if and only if (1) } b = \beta \text{ is in } L, \text{ and } B \in \text{Exp}(P[A]), \text{ or (2) } \beta \text{ is in } LC, \text{ and } B \in \text{Sup}(P[A](X|Y)) = \text{Sup}(P(X|Y \cap A)).$$

Of course, the expectation test is indistinguishable from the Ramsey test when countable additivity is enforced. This means that the fact that $P(A|B) = 1$ is in this case perfectly represented by the fact that $B$ is expected (for $P$) given $A$. But, once we consider finitely additive models, $P(A|B) = 1$ is only a necessary but not a sufficient condition for $B$ being expected given $A$ (for $P$). It is worth pointing out that the expectation test might have its own separate interest. In fact, the same rationale that motivates the use of the Ramsey test for probability conditionals motivates the basic idea behind the test.69 As before, when $F(P) \cap A = \emptyset$, we set $\text{Exp}(P[A])$ to $\emptyset$. By the same token, the expectation of the abnormal two-place function is set to $\emptyset$.

The export-import laws are a rather robust part of the probabilistic program. In fact, they continue to be validated under either test. A final

---

68 van Fraassen did appeal to this formal maneuver in order to define full belief in “Fine-grained Opinion, Probability, and the Logic of Full Belief.”

69 The exact formal details of the conditional logics induced by the expectation test (and by the generalized test) remain an open problem in the field.
issue remains open: Is there a new qualitative condition representing $P(A|B) = 1$ for finitely additive models? The answer is “yes.” In fact, for non-coreless functions $P$ we can define:

\[ K \text{ is a kernel of a core } C \text{ for function } P \text{ if and only if } K \subseteq C \text{ and } P(K|C) = 1 \]

So, for non-coreless $P$, we can define the probabilistic kernel for $P$, as the set of all propositions entailed by some kernel of some core for $P$:

\[ PK(P) = \{ A : A \text{ is entailed by some kernel of some core of } P \} \]

Observation (2): if the universe $U$ of the space $S$ is normal for a function $P$ defined on $S$, and $A$ is epistemically possible for $P$, $P(B|A) = 1$ if and only if $B \in PK(P[A])$.\(^{70}\)

Although we can easily define acceptance conditions in terms of what is entailed by probabilistic kernels, the idea seems to have less intuitive interest than the one based on the expectation test. A curious situation thus arises where an acceptance test, which formally departs from Adams’s test (the expectation test), might, however, preserve in a better manner the intuitions that motivate probabilistic acceptance tests on qualitative grounds—that is, in spaces of at most countable size. The complete formal details of the conditional systems validated by each test are still unknown. My main point in this section was to stress the fact that export-import is a salient part of the program in probabilistic semantics, even when we consider large classes of models and important weakenings of the axiomatic base used to characterize both conditional and unconditional probability.

HORACIO ARLO-COSTA

Carnegie Mellon University

\(^{70}\) Observation (2) can be easily established as follows: assume first $B \in PK(P[A])$. Therefore, there is a core $C$ such that it has a kernel $K$ entailing $B$. So, $P[A](K) = 1 = P(K|A)$. Finite additivity and the inferred normality $A$ gives us then that $P(B|A) = 1$. Assume now that $P(B|A) = 1$. We have to consider two main cases. First assume that $B$ does not cut any core $C$ for $P[A]$. This leads to a contradiction because $P[A](C|U) = 1$ for an arbitrary core for $P$. Therefore, $P[A](C + B|U) = 2$ by finite additivity and the assumed normality of $U$. Assume now that there is a core $C$ for $P[A]$ such that $C \cap B \neq \emptyset$. Then, $C \cap B$ carries probability 1 according to $P[A]$. For assume that there is $Z$ in $C$ such that it does not cut $C \cap B$ and $P[A](Z|U) > 0$. Then, by finite additivity, $P[A](B + Z|U) = P[A](B|U) + P[A](Z|U) > 1$, taking into account that we assumed $P[A](B|U) = 1$. Since $P[A](C|U) = 1$, axiom II guarantees that $C \cap B$ is a kernel for $C$ for the function $P[A]$, and this kernel entails $B$. 