ALAN HÁJEK

WHAT CONDITIONAL PROBABILITY COULD NOT BE

ABSTRACT. Kolmogorov's axiomatization of probability includes the familiar ratio formula for conditional probability:

$$\text{(RATIO)} \quad P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad (P(B) > 0).$$

Call this the *ratio analysis* of conditional probability. It has become so entrenched that it is often referred to as the *definition* of conditional probability. I argue that it is not even an adequate analysis of that concept. I prove what I call the *Four Horn theorem*, concluding that every probability assignment has uncountably many 'trouble spots'. Trouble spots come in four varieties: assignments of *zero* to genuine possibilities; assignments of *infinitesimals* to such possibilities; *vague* assignments to such possibilities; and *no* assignment whatsoever to such possibilities. Each sort of trouble spot can create serious problems for the ratio analysis. I marshal many examples from scientific and philosophical practice against the ratio analysis. I conclude more positively: we should reverse the traditional direction of analysis. Conditional probability should be taken as the primitive notion, and unconditional probability should be analyzed in terms of it.

"I'd probably be famous now
If I wasn't such a good waitress."
Jane Siberry, "Waitress"

Over three hundred years ago, Fermat and Pascal began the enterprise of systematizing our pre-mathematical, pre-philosophical concept of probability. Orthodoxy has it that this task was completed in 1933, when Kolmogorov provided his axiomatization. In the process, he systematized another pre-theoretical concept: *conditional probability*. He identified conditional probability with a certain ratio of unconditional probabilities, according to the formula:

$$\text{(RATIO)} \quad P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad (P(B) > 0).$$

I will call this the *ratio analysis* of conditional probability.[1] It has become so entrenched that we seem to have forgotten that it *is* an analysis. My main purposes in this paper are to cast doubt on its adequacy, and to argue

Synthese **137:** 273–323, 2003.

that conditional probability should be taken as the fundamental notion in probability theory.

1. THE RATIO IS AN ANALYSIS, NOT A DEFINITION, OF CONDITIONAL PROBABILITY

What I have just said may strike you as absurd. You may say that conditional probability is just a technical term that is *defined* by this formula. If so, you are in good company. For example, Jackson (1987, p. 12) writes: "it is important to take the conditional probability of consequent given antecedent to be *defined* as the probability of the conjunction of the antecedent and the consequent divided by the probability of the antecedent". (The emphasis is his.) Skyrms (1980, p. 156; 2000, p. 119) and Earman (1992, p. 36) similarly speak of (RATIO) as the "definition" of conditional probability, just to mention two other prominent authors, among many. It is quite understandable that they speak this way – it is part of the orthodoxy, after all. And it only underscores, to my mind, the extent to which the ratio analysis has become dogma.

To be sure, *if* "the conditional probability of A given B" were a purely technical notion, then Jackson, Skyrms, Earman, etc. would doubtless be right.[2] There is nothing problematic about introducing the symbol '|' into the language of the probability calculus, and defining it in terms of the ratio formula. You can even intone the words "the conditional probability of A given B" when reading '$P(A \mid B)$' out loud, if you insist. If these words are not loaded with philosophical and commonsensical associations, then this is no more contentious than saying "the probability of A slash B", and taking that as convenient shorthand for the ratio.

But the words are loaded. They are tinged with associations with models of learning, rational belief revision and decision-making, with analyses of confirmation, causal dependence, the semantics of conditionals, and so on. Jackson goes on to deploy intuitions about conditional probabilities in his defense of Adams' Thesis that the assertability of the indicative conditional $A \rightarrow B$ goes by the conditional probability of B, given A. For example, he writes: "Take a conditional which is highly assertible, say, 'If unemployment drops sharply, the unions will be pleased'; it will invariably be one whose consequent is highly probable given the antecedent. And, indeed, the probability that the unions will be pleased given unemployment drops sharply is very high" (12). Likewise, Skyrms goes on to deploy intuitions about conditional probabilities in his defense of his account of laws as resilient generalizations; and Earman goes on to deploy intuitions about conditional probabilities in his defense of the Bayesian analysis of confirmation. They are clearly appealing to some commonsensical notion (albeit refined by philosophical reflection) rather than to something that is merely stipulatively defined.

After all, we had the concept of conditional probability long before we donned our philosopher's or mathematician's caps. In ordinary language, conditional probability statements can be made using locutions of the form "it's likely that p, given q", "it's improbable that r, if s", and variations on these. In *The Piano Man*, Billy Joel sings, "I'm sure that I could be a movie star, if I could get out of this place": his probability for becoming a movie star is high, not unconditionally, but conditionally on his getting out of this place. And so forth. Conditional probability is not just a technical notion such as 'zero-sum game' or 'categorical imperative' or 'Turing computability'. Rather, it is a familiar concept, like 'game', or 'moral duty', and I daresay more familiar than 'computability'. It is there to be analyzed, if possible. (If it is not possible, then it follows immediately that the ratio does not provide an adequate analysis of it.) Our choice of words in reading $P(A \mid B)$ as 'the probability of A, given B' is not so innocent after all. And far from being a stipulative definition of that concept, I will argue that the ratio formula is not even an adequate analysis of it.[3]

In the sequel, I will follow convention in using the notation '$P(A \mid B)$' as shorthand for the ratio $P(A \cap B)/P(B)$. However, I want you to resist the further convention of reading it as 'the probability of A given B'. When I want to speak of the conditional probability of A, given B, while remaining neutral as to how this might be analyzed (if at all), I will simply write '$P(A$, given $B)$'. I will call B 'the condition'.

My strategy will be as follows. After some stage-setting in Section 2, I will prove in Section 3 what I call the *Four Horn theorem*. Its upshot is that every probability assignment has uncountably many 'trouble spots'. Trouble spots come in four varieties: assignments of *zero* to genuine possibilities; assignments of *infinitesimals* to such possibilities; *vague* assignments to such possibilities; and *no* assignment whatsoever to such possibilities. The next four sections show how each sort of trouble spot can create serious problems for the ratio analysis. Section 4 rehearses, then gives a new gloss on, an old problem: that the analysis is mute whenever the condition has probability zero, and yet conditional probabilities may nevertheless be well defined in such cases. Section 5 discusses problems for infinitesimal assignments to the unconditional probabilities that figure in (RATIO). Section 6 shows that the analysis cannot respect the fact that various conditional probabilities are sharp even when the corresponding unconditional probabilities are *vague*. Section 7 argues that various conditional probabilities are defined even when the corresponding unconditional probabilities are undefined, and indeed *cannot* be defined. In Section 8, I canvass many instances from actual scientific and philosophical practice. The dogma has it that the ratio is an adequate analysis of 'the probability of A, given B', irrespective of the interpretation of probability itself. I give arguments against the ratio analysis for both subjectivist and objectivist construals of probability.

If (RATIO) fails as an analysis of conditional probability, how can we explain its undoubted utility? In Section 9, I argue that it provides a *constraint* on conditional probability: when $P(A \cap B)$ and $P(B)$ are both sharply defined, and $P(B)$ is non-zero, the probability of A given B is constrained to be their ratio. In the cases that I discuss, these conditions are not met, so there is no constraining to be done by the ratio. I conclude more positively in Section 10: we should reverse the traditional direction of analysis. Conditional probability should be taken as the primitive notion, and unconditional probability should be analyzed in terms of it.

2. BACKGROUND

There is some philosophical and technical background that I will be assuming in my arguments.

2.1. *Subjective/Objective Probability*

I assume that probability comes in at least two varieties: subjective and objective. I will stay neutral as to how exactly these should be understood; suffice to say that my examples will include putatively paradigm cases of each. The following characterizations, though not fully fleshed-out and in some ways problematic (topics for other occasions), will be more than adequate for my purposes here:[4]

Subjective probability: Probability is interpreted as the degree of belief, or *credence*, of a rational agent. Let the agent be you, for example. There are various proposals for attributing a single probability function P to you. For example, de Finetti (1980) regards P as capturing all information about the bets that you are prepared to enter into: $P(X)$ is the price that you regard as fair for a bet that pays 1 unit if X occurs, and 0 otherwise. According to van Fraassen (1990), P encapsulates all information about the judgments that you make – for example, if you judge it as likely as not that Madonna will win a Nobel prize, then P(Madonna will win a Nobel prize) = 1/2. Ramsey (1926), Savage (1954) and Jeffrey (1983b) derive both probabilities and utilities (desirabilities) from rational preferences (given various assumptions about the richness of the preference space, and certain "consistency" assumptions on the preference relation). In a similar spirit, Lewis (1986a, 1994a) analyses P as the probability function belonging to the utility function/probability function pair that best rationalizes your behavioral dispositions (rationality being given a decision-theoretic gloss).

Objective probability: This is often called objective *chance*. It exists heedless of our beliefs and interests. Long run relative frequency is typically a good guide to determining it. (Some, e.g., Reichenbach 1949, think that, more than being a good guide, such relative frequency should be *identified* with objective chance; but I part company with them there – see my 1997.) Chance also places constraints on rational credence, via something along the lines of Lewis' (1980, 1986b) *Principal Principle* – more on that in Section 8.5. Some authors (e.g., Popper (1959a), Mellor (1971), Fetzer (1983), and many others – see Gillies (2000) for a survey) regard it as a kind of propensity, or disposition, or causal tendency for an outcome to be brought about by an experimental arrangement. For example, according to Popper, a probability p of an outcome of a certain type is a propensity of a repeatable experiment to produce outcomes of that type with limiting relative frequency p. Lewis among others contends that chances evolve over time, the past (relative to any given moment of time) always having a chance of 1. Quantum mechanical probabilities are standardly regarded as objective. Chance values are plausibly determined by the laws of nature. Some authors (e.g., Lewis 1986b, p. 118) regard a deterministic world as one in which all chances are 0 or 1.

2.2. *Vague Probability*

Let us drop the assumption that a given proposition is assigned a unique, sharp value, and allow instead that it may be assigned a range of values. I will follow van Fraassen's (1990) representation of vague subjective probability (which develops proposals by Levi (1974, 1980) and Jeffrey (1983a), among others). I will then extend it to the case of objective probability.

Suppose that your state of opinion does not determine a single probability function, but rather is consistent with a multiplicity of such functions. In that case, your opinion is represented as the *set* of all these functions. Call this set your *representor*. Each function in your representor corresponds to a way of precisifying your opinion in a legitimate way. Imagine, for example, that your probability for Diego Maradonna winning an Oscar is not a sharp value; rather, it lies in some interval, say [1/2, 3/4]. Then your representor contains all probability functions that assign a sharp value in the interval [1/2, 3/4] to Maradonna's winning, and that are otherwise compatible with your state of mind. This might be thought of as a probabilistic analogue of the supervaluational approach to vagueness.

More controversially, let me suggest that we remain open to the possibility of vague *objective probabilities*. For example, perhaps the laws of nature could be vague, and if these laws are indeterministic, then objective probabilities could inherit this vagueness.[5] Be that as it may, I submit that vague objective probabilities, if there are any, could likewise be understood in terms of representors. For example, a chance that is vague over the set S corresponds to a set of sharp chances, taking on each of the values in S.

2.3. *Probability Gaps*

Furthermore, we can make sense of probability *gaps* – propositions that get assigned no probability value whatsoever – under both the subjective and objective interpretations.

As a point about probability functions in the abstract, this is familiar enough. We begin with a probability space $(\Omega, \mathcal{F}, P)$, where Ω is some set of 'elementary events', or 'possible worlds', and $\mathcal{F}$ is a sigma-algebra of subsets of Ω. P is defined on all the members of $\mathcal{F}$, but on no other propositions. So any subset of Ω that does not belong to $\mathcal{F}$ can be thought of as a probability gap as far as P is concerned: a proposition to which P assigns no value.

Decision theory recognizes the possibility of probability gaps in its distinction between *decisions under risk*, and *decisions under uncertainty*: in the latter case, probabilities are simply not assigned to the relevant states of the world. Statistics recognizes such gaps in its distinction between *random* and *unknown* parameters: probabilities are not assigned to values of the latter. And indeed it seems that we must allow for probability gaps on all of the accounts of subjective probability given above. On the betting account, this is clear: suppose that you are currently not prepared to enter into *any* bet regarding X – for example, you have never even considered X. (Of course, you might *later* become prepared to enter into such bets, having giving X some thought, or upon coercion.) Likewise on the preference-based accounts, for your preferences may not be defined over all propositions. Probability gaps are a little harder to come by on van Fraassen's and Lewis' accounts: if you lack determinate judgments concerning X, or if you lack the relevant behavioral dispositions, then it seems that all precisifications of the probability of X are legitimate, so that your probability for X is vague over the entire [0, 1] interval – and that seems to be something different from a probability gap. But suppose that you resolutely *refuse* to assign X a probability, or for some reason *cannot* assign X a probability. Then any precisification of the probability of X

seems to distort your opinion, giving it a value where you choose to remain silent. In that case, your opinion would be better represented as having a gap for X. Perhaps, for example, you regard X as lacking a chance value altogether, and you think that rational opinion concerning X should follow suit. Or perhaps X involves a future action of yours upon which you are deliberating; more on this in Section 8.4.

Or perhaps you regard X as *non-measurable* – a set that simply cannot be assigned a probability, consistent with certain natural constraints that one would want to impose. Non-measurable sets notoriously arise in interesting and surprising ways. For instance, consider the case of assigning a 'uniform' probability distribution P across the [0, 1] interval. Certain subsets of the interval simply cannot be assigned any probability whatsoever by P, and so are non-measurable (with respect to it). It is not that they receive zero probability; rather, they are probability gaps. We may prove this, à la the textbook, as follows:

By Kolmogorov's axiomatization, P is countably additive: for every infinite sequence of mutually disjoint propositions $A_1, A_2, \ldots$ in its domain,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Assume further that it is uniform in the sense of being translation invariant: for any real number Δ, and any set X, $P(X) = P(X + \Delta)$, where $X + \Delta$ is the set produced by adding Δ (modulo 1) to each member of X. Consider the equivalence relation on [0, 1] satisfied by x and y iff $x - y$ is rational. This partitions [0, 1] into equivalence classes. Let C be a choice set containing exactly one member of each of these classes (here we assume the axiom of choice). For each rational r in [0, 1], let C_r be the set you get by adding (modulo 1) r to each member of C. There are denumerably many C_r's, and they are mutually exclusive. Their union is [0, 1]. Suppose for *reductio* that each C_r receives some probability value; by the assumption of uniformity every C_r receives this same value. Then either this value is 0, in which case $\sum P(C_r) = 0$, or it is some positive number, in which case $\sum P(C_r) = \infty$. By countable additivity, we have either 1 = 0 or 1 = ∞. Contradiction. This completes the *reductio*. We cannot maintain the assumption that each C_r receives some probability value – but if any one of them doesn't, then they all don't (by the assumption of uniformity). All the C_r's are thus non-measurable sets. That is, we have proven the existence of infinitely many probability gaps in the [0, 1] interval – a surprising and profound result that I will appeal to later on. If we can interpret 'P' both subjectively or objectively, we have both subjective and objective probability gaps – indeed, uncountably many.

2.4. *Infinitesimal Probabilities, Countable Additivity, and Regularity*

Robinson (1966) exploited compactness to show that there are non-standard models of first order analysis containing *positive infinitesimal* elements: positive numbers smaller than every positive real. Consider a first order language appropriate for real analysis – in particular, for each real number r, there is a name $\mathbf{r}$. The theory of analysis consists of all the sentences of this language that are true according to the standard model. Consider the theory that is the union of the theory of analysis, with all sentences of the form $\mathbf{r} < y$, for each real r. Each finite subset of this theory has a model; hence, by compactness, this theory has a model as well. Any such model (and there are many) is a non-standard model of the real numbers. The denotation of y in such a model is an infinite element, greater than each of the standard reals. Its reciprocal $1/y$ is a positive infinitesimal element.[6] In what follows I will drop the word 'positive' before 'infinitesimal': while 0 is strictly speaking an infinitesimal (the only real one, in fact), my interest in infinitesimals will concern the positive ones, and when philosophers speak of infinitesimal probabilities, it's typically the positive ones they have in mind.

Kolmogorov's axiomatization stipulates that probability measures are real-valued; hence according to that axiomatization, there can be no infinitesimal probabilities. We may wish to relax the stipulation, and indeed philosophers such as Skyrms (1980) and Lewis (1980) explicitly embrace infinitesimal probabilities. However, the stipulation is not idle. For as I have said, Kolmogorov also required probabilities to be *countably additive*. But the notion of countable additivity does not even make sense in the context of non-standard analysis, and there is no obvious analogue of it. For the very statement of countable additivity requires the taking of a 'limit as n tends to infinity' of a sequence of partial sums of probabilities, and it turns out that this is not something that can even be said in the language of non-standard analysis (the set of natural numbers, which gets appealed to in the ϵ-N definition of the limit, is not an 'internal' set).

Finally, a probability function is said to be *regular* iff it assigns probability 1 only to logical truths, and 0 only to contradictions. (For our purposes, we might equally define a regular function as one that assigns 1 only to the necessary proposition(s), and 0 only to the impossible proposition(s) – any differences that there might be between the two definitions will not matter to us here, and I will use both sentential and propositional locutions.) A number of authors propose versions of the idea that rational credence functions are regular (sometimes calling them ‘strictly coherent’), among them Kemeny (1955), Jeffreys (1961), Carnap (1963), Edwards et al. (1963), Stalnaker (1970), Shimony (1970), and Appiah (1985). Note that an agent who updates by conditionalization upon some item E of evidence ends up with an irregular probability function, for $P(E)$ becomes 1. So the proposal of regularity contradicts another widely endorsed claim about rationality: that updating should take place by conditionalization. For further opposition to the demand for regularity, see de Finetti (1972) and Levi (1978).[7]

We turn now to the promised Four Horn theorem. Each of the four sections that follow it (Section 4–7) focuses on a type of ‘trouble spot’ for the ratio analysis: zero probabilities, infinitesimal probabilities, vague probabilities, and undefined probabilities. The discussion in these sections is largely self-contained, and the reader who is impatient with technicalities (or simply impatient) may want to skip directly to them. What the theorem, coupled with a widely accepted assumption, show is that the trouble spots are inescapable, and that they are, to put it mildly, plentiful. There’s trouble, and a lot of it.

3. THE FOUR HORN THEOREM[8]

We begin by proving a theorem concerning probability measures (that need not be total functions). As usual, they are real-valued, and sharp (that is, single functions). We then relax these assumptions, generalizing the theorem for probability functions that may also take infinitesimal values, giving a corollary for such functions that are regular, and then a theorem concerning vague probability assignments. This culminates in the Four Horn theorem, which puts the results together. This theorem, combined with an assumption regarding propositions, implies that every probability function has uncountably many trouble spots. These trouble spots, moreover, will be central to my attack on the ratio analysis.

PROBABILITY MEASURE THEOREM. Let Ω be an uncountable set. Any (real-valued, sharp) probability measure fails to assign positive probability to uncountably many subsets of Ω.

Proof. Let uncountable Ω be given, and let P be a (real-valued, sharp) probability measure. Let D be an uncountable set of disjoint subsets of Ω. (The disjointness will come in handy when we appeal to finite additivity.) For each $n = 1, 2, \ldots$ consider

$$G_n = \{S \in \mathrm{D}\colon P(S) \geq 1/n\}.$$

Each set in G_1 has probability at least 1; it thus has at most 1 member.
Each set in G_2 has probability at least 1/2; it thus has at most 2 members (by finite additivity of P) ...
For each i, G_i has at most i members (by finite additivity of P), and is thus a finite set.

Therefore, $\cup_i G_i$ is countable, being a countable union of finite sets. This is the set of members of D that receive positive, non-infinitesimal assignments from P (since for any positive, non-infinitesimal x, there is an n such that $x \geq 1/n$). This leaves uncountably many members of D, that is, subsets of Ω, that do *not* receive positive, non-infinitesimal assignments from P.

Since P is real-valued, it does not assign positive, infinitesimal probabilities; thus, the word 'non-infinitesimal' was redundant in the previous sentence. That is, we have shown that uncountably many subsets of Ω do not receive positive assignments from P. Q.E.D.

How can a (real-valued, sharp) probability measure P fail to assign positive, non-infinitesimal probability to a given set? There are two ways: by assigning it 0 probability, or by assigning it no probability whatsoever. Now let us relax the assumption that P is a measure, allowing it to take infinitesimal values; in that case we will still call P a 'probability function', even though strictly speaking we have gone beyond Kolmogorov's axiomatization. Then there is a third way in which a probability function P can fail to assign positive, non-infinitesimal probability to a given set: by assigning it infinitesimal probability. We have a strengthening of the probability measure theorem:

PROBABILITY FUNCTION THEOREM (infinitesimals allowed). Let Ω be an uncountable set. Any (sharp) probability function either assigns 0 to uncountably many subsets of Ω, or assigns infinitesimal probability to uncountably many subsets of Ω, or fails to assign probability to uncountably many subsets of Ω.

Proof. The proof is as before, up to and including: "This leaves uncountably many members of D, that is, subsets of Ω, that do not receive positive, non-infinitesimal assignments from P". Then we reason as follows. The following union of sets is uncountable:

$$\{S \in \mathrm{D} : P(S) = 0\} \cup \{S \in \mathrm{D} : P(S) \text{ is infinitesimal}\} \cup \{S \in \mathrm{D} : P(S) \text{ is undefined}\}.$$

Thus, at least one of these three sets of subsets of Ω is uncountable (for if each of them were countable, their union would be also). Q.E.D.

With minimal extra work, we get:

REGULARITY THEOREM. Any regular probability function defined on uncountably many mutually disjoint propositions assigns infinitesimal probabilities to uncountably many propositions.

Proof. Let P be a regular probability function defined on an uncountable set D of mutually disjoint propositions (and on other sets, too, of course). The proof goes through as before. Then note that since P is regular, it assigns 0 only to the empty set. Thus $\{S \in \text{D}: P(S) = 0\}$ cannot be uncountable; indeed, it has at most one member. Nor can $\{S \in \text{D}: P(S)$ is undefined$\}$ be uncountable; indeed, it is empty (by definition of D). This leaves $\{S \in \text{D}: P(S)$ is infinitesimal$\}$ as the set that must be uncountable. Q.E.D.

So far, our theorems have concerned (sharp) single probability functions. Vague probability assignments, on the other hand, are represented by (non-singleton) sets of such functions, or representors. So now let the term 'probability assignment' subsume both sharp probability functions and vague probability assignments.

VAGUE PROBABILITY THEOREM. If a probability assignment defined on an uncountable algebra of propositions is vague at all, then there are uncountably many propositions to which it assigns vague probability.

Proof. Let a vague probability assignment V defined on an uncountable algebra of propositions $\mathcal{U}$ be given. It corresponds to a representor with at least two probability functions defined on $\mathcal{U}$. Choose two such functions, P_1 and P_2. V assigns vague probability to each proposition on which they differ. Hence it suffices to show that they differ on uncountably many propositions.

Since P_1 and P_2 are distinct, there is some proposition $D \in \mathcal{U}$ on which they differ:

$$P_1(D) \neq P_2(D).$$

Thus, $P_1(D^c) \neq P_2(D^c)$, where $D^c \in \mathcal{U}$ also.

Since $\mathcal{U}$ is uncountable, either

(i) there are uncountably many propositions in $\mathcal{U}$ disjoint from D, or
(ii) there are uncountably many propositions in $\mathcal{U}$ disjoint from D^c.

(Any proposition $X \in \mathcal{U}$ that has non-empty intersection with both D and D^c can be split into two parts: $X \cap D \in \mathcal{U}$, which is disjoint from D^c, and $X \cap D^c \in \mathcal{U}$, which is disjoint from D.) Without loss of generality, suppose (i) is the case. If P_1 and P_2 differ on uncountably many of these propositions, we are done. So suppose that P_1 and P_2 differ on only countably many of these propositions; they thus *agree* on uncountably many of them. But for each such proposition A on which they agree, they disagree on $D \cup A$:

$$\begin{aligned} P_1(D \cup A) &= P_1(D) + P_1(A), \text{ since } D \text{ and } A \text{ are disjoint,} \\ &= P_1(D) + P_2(A), \text{ since } P_1 \text{ and } P_2 \text{ agree on } A, \\ &\neq P_2(D) + P_2(A), \text{ since } P_1 \text{ and } P_2 \text{ differ on } D, \\ &= P_2(D \cup A), \text{ since } D \text{ and } A \text{ are disjoint.} \end{aligned}$$

Each of the uncountably many candidates for $D \cup A$ is a proposition for which P_1 and P_2 differ; each of them is thus a proposition to which V assigns vague probability. Q.E.D.

Every probability assignment is either sharp or vague. Putting our results together, we have the

FOUR HOUR THEOREM. Any probability assignment defined on an uncountable algebra on an uncountable set either

1. assigns zero probability to uncountably many propositions; or
2. assigns infinitesimal probability to uncountably many propositions; or
3. assigns no probability whatsoever to uncountably many propositions; or
4. assigns vague probability to uncountably many propositions

(where the 'or's are inclusive).

We now make an assumption about propositions, one that many will regard as not just true, but necessarily so.

ASSUMPTION. There are uncountably many mutually disjoint propositions.

JUSTIFICATION. For each real number in [0, 1], there corresponds a distinct proposition. For example, imagine throwing an infinitely fine dart at the interval. For each r in the interval, there is a proposition L_r to the effect that the dart lands exactly on r. Moreover, these propositions are mutually disjoint. If you prefer, consider instead a particular radium atom; for each time $t > 0$ there is a proposition that the atom decays exactly at time t, disjoint with all the others. Or if you prefer to think in terms of possible worlds (be they concrete, abstract, or what have you), and propositions as sets of possible worlds: there are uncountably many possible worlds, thus infinitely many mutually disjoint propositions (e.g., the singleton sets of these worlds; cf. Lewis 1986c).

We are now in the realm of metaphysics rather than mathematics. Combining our mathematical theorem with our metaphysical assumption, we have the more startling result: *any probability assignment whatsoever has uncountably many trouble spots: propositions that it assigns zero probability, infinitesimal probability, vague probability, or no probability at all.* If it is defined on an uncountable algebra of propositions of an uncountable set, the mathematics alone does the work. If it is not, the metaphysics chimes in: whatever algebra it *is* defined on, *outside* that algebra it is undefined on uncountably many propositions.

The following four sections address the four sorts of trouble spots for the ratio analysis that we find in the Four Horn theorem. It will prove convenient to take them slightly out of order: zero probability assignments (Section 4), infinitesimal probability assignments (Section 5), vague probability assignments (Section 6), and no probability assignment whatsoever (Section 7). It is inescapable that any probability assignment lands on at least one of the four horns. Moreover, it is noteworthy that each horn is actually *occupied* by various interesting probability functions. In each section I will give examples that are both paradigm cases of what it is like to land on the relevant horn and of independent philosophical interest.

4. THE ZERO DENOMINATOR PROBLEM

Suppose that we have a probability assignment that lands on the first horn: it assigns zero probability to uncountably many propositions. Then for each such proposition, all probabilities conditional on it are undefined as far as the ratio analysis is concerned. After all, (RATIO) comes with a proviso: the condition must have positive probability. Jackson remarks: "The restriction to cases where $P(B) > 0$ is less severe than may at first appear" (11, adapting his notation). His argument for this is essentially that all rational credence functions are regular. Again, I am only using Jackson as a foil, and this opinion is widely voiced. But I believe that on the contrary, the restriction is *more* severe than may at first appear. The ratio analysis fails to give values to certain conditional probabilities that apparently are well-defined. Let us call this *the zero denominator problem*.

The problem is well known.[9] Indeed, given how many probability textbooks go out of their way to caution the reader that 'probability 0 does not imply impossible', it is perhaps surprising that more is not made of it, at least in philosophical circles. In any case, let me mention some particularly interesting cases in which it arises – some familiar, some less so.

4.1. *Some Textbook Examples*

Continuous random variables give rise to non-trivial events of probability 0. Associated with such a random variable is a probability density function (think, for example, of the 'bell'-shaped normal distribution). The probability of a continuous random variable X, taking a particular value x, equals 0, being the integral from x to x of X's density function. Yet it would seem that various probabilities, conditional on this event, are well-defined, for example:

$$P(X = x,\ \text{given } X = x) = 1.$$

Indeed, here and throughout the paper, we hold this truth to be self-evident: the conditional probability of any (non-empty) proposition, given itself, is 1. I think that this is about as basic a fact about conditional probability as there can be, and I would consider giving it up to be a desperate last resort. Some others are comparably basic; in the case before us they include:

> $P(X \neq x,\ \text{given } X = x) = 0$,
>
> $P(\mathbf{T},\ \text{given } X = x) = 1$, where $\mathbf{T}$ is a necessarily true proposition (e.g., "everything is self-identical"),
>
> $P(\mathbf{F},\ \text{given } X = x) = 0$, where $\mathbf{F}$ is a necessarily false proposition (e.g., not-$\mathbf{T}$).

Less trivially, various conditional probability assignments based on judgments of independence are compelling – for instance:

$$P(\text{this coin lands heads, given } X = x) = 1/2.$$

Further examples are provided by the behavior of relative frequencies of events in independent, identically distributed trials. Various laws of large numbers concern the convergence of such relative frequencies to the corresponding true probabilities. However, they do not *guarantee* that relative frequencies will so converge; rather, the convergence only occurs 'with probability 1'. This caveat isn't just coyness – a fair coin *could* land heads every toss, forever, although its doing so is an event of probability 0. If you think that this sequence of results cannot happen 'because' it has probability 0, then you should think that no sequence can happen, since each sequence has the same probability. Again, however, various probabilities conditional on such events would seem to be well defined: P(the coin lands heads forever, given the coin lands heads forever) = 1, and so on.

We may regard the probabilities in these cases objectively or subjectively. Other cases arise specifically for objective probabilities.

4.2. *Chances*

It would appear that objective probability functions are not regular. For example, it is plausible that the *chance* function P_t at any given time t lands on the first horn of the Four Horn theorem. Lewis, for example, contends that the past is not chancy: any proposition whose truth value is settled at a time earlier than t is assigned either 1 (if it is true), or 0 (if it is false) by P_t. And there are uncountably many such propositions that are false. Suppose that a coin landed heads at 12:00, and consider

$P_{12:01}$ (the coin landed tails, given the coin landed tails).

This conditional probability should equal 1, but according to the ratio formula, it is undefined. Similar points can be made using each of the uncountably many false propositions about the past as the condition.

Or consider some proposition about the *future* – for example, 'an eclipse will occur at 15:52', and the conditional probability:

$P_{12:01}$ (an eclipse will occur at 15:52, given an eclipse will occur at 15:52).

This should equal 1, whether or not the condition is false. And according to Lewis among others, if the world is deterministic, then all false propositions have chance 0 at any time. In that case, if the condition is false, this conditional probability is undefined according to the ratio analysis; if the condition is true, change the example, replacing 'occur' with 'not occur'. Thus, the adequacy of the ratio analysis is held hostage to empirical matters: the world had better not be deterministic! Clearly, the ratio analysis cannot have empirical consequences. So if the conditional probability is 1, irrespective of whether the world is deterministic, the analysis is refuted. Similar points can be made using conditional probabilities involving **T** and **F**, as before.

The proponent of the ratio analysis of conditional chance is thus committed to saying either

(1) the chance function (at any given moment of time) is *regular*; or
(2) conditional chances of the form $P(X$, given $Z)$ are never defined when $P(Z) = 0$.

(1) is implausible; logic and chance surely do not align so neatly. In any case, (1) appears to be an empirical claim, one hardly discoverable by pure conceptual analysis. But (2) is hardly appealing, either, as I have argued. Indeed, conjoined with the claim that the chance function is irregular at some time, it contradicts a truth we hold to be self-evident.

4.3. *Subjective Probabilities*

Still other cases arise more naturally for subjective probabilities. Let P be the credence function of a rational agent. If, *contra* Kemeny, Jeffreys et al., P is irregular, then it assigns 0 to at least one logically possible proposition. Let Z be such a proposition. I claim that rationality requires the following assignments: $P(Z$, given $Z) = 1$, $P(Z^c$, given $Z) = 0$, $P(\mathbf{T}$, given $Z) = 1$, and $P(\mathbf{F}$, given $Z) = 0$. But an advocate of the ratio analysis cannot even grant that rationality *permits* these assignments.

We have already noted that updating by conditionalization leads one to irregularity. Suppose, for example, that you begin by assigning probability 1/2 to a given die landing with an odd face showing up. The die is tossed, and you conditionalize on the evidence that the die landed odd. You then assign probability 0 to the die landing with an even face showing up on that toss. Still, surely rationality permits you to assign P(die landed even, given die landed even) = 1 (and, I would add, even requires it). We have, then, a rational credence function P that assigns a conditional probability that cannot be equated to a corresponding ratio of unconditional probabilities.

There are uncountably many propositions on which conditionalization could take place. It seems, then, that there are uncountably many candidates for Z for some rational credence functions. They thus land on the first horn of the Four Horn theorem. But that is just icing on the cake – one such candidate is all I need to make my point.

We have a dilemma. Either:

(1) all rational credence functions are regular; or
(2) for all rational credence functions P, conditional probabilities of the form $P(X$, given $Z)$ are never defined when $P(Z) = 0$.

As we have seen, various authors opt for (1). They thus commit themselves to saying that updating by conditionalization is *always* irrational. They also seem to have a novel 'proof' of the non-existence of God (as traditionally conceived) – for an omniscient agent who is certain of which world is the actual world is thereby convicted of irrationality! By their lights, rationality requires a certain form of ignorance regarding all contingent propositions, or else false modesty. And even if regularity is such a requirement on rationality, we all fall short of such requirements. But when we (irrationally?) assign 0 to a contingent proposition A, we still seem to have no trouble giving various corresponding conditional probability assignments that are apparently rationally required, and surely rationally permitted – for example, $P(A$, given $A) = 1$.[10] (2) may seem to be more palatable, then. But do we really want to say that rationality does not even *permit* conditional probabilities such as $P(Z$, given $Z)$ to be defined?

4.4. *Symmetric Distributions*

Symmetric probability assignments over uncountable sets give rise to particularly acute problems for the ratio analysis, because many (indeed, uncountably many) non-trivial conditional probabilities with probability zero conditions are apparently well-defined, and their values are obvious. Here are some familiar examples.

(i) (Due to Borel.) Suppose that we have a uniform probability measure over the Earth's surface (imagine it to be a perfect sphere). What is the probability that a randomly chosen point lies in the western hemisphere (W), given that it lies on the equator (E)? 1/2, surely. But the probability that the point lies on the equator is 0, since the equator has no area.

And the problem could be made to go the other way. We could embed the Earth's surface in a much larger space, relative to which it has measure 0. Yet P(the point lies on the earth's surface, given the point lies on the earth's surface) = 1, and still P(the point lies in the western hemisphere, given the point lies on the equator) = 1/2.

(ii) Imagine throwing an infinitely fine dart at the [0,1] interval. Suppose that the probability measure for where the dart lands is uniform over the interval – the so-called 'Lebesgue measure'. What is the probability that the dart lands on the point 1/4, given that it lands on either 1/4 or 3/4? 1/2, surely. But the probability that the point lands on 1/4 or 3/4 is 0 according to the uniform measure.

The probability functions associated with these examples land on the first horn of the Four Horn theorem: they assign probability 0 to uncountably many propositions. In (i), we could have asked a similar question for each of the uncountably many lines of latitude, with the same answer in each case. In (ii), we could have asked a similar question for each of uncountably many pairs of points in [0, 1], with the same answer in each case. In all of these cases, the ratio analysis fails to yield an answer for the conditional probability, yet it is seemingly well-defined in each case.

We could interpret the uniform probability here either objectively or subjectively. In (i), there might be some genuinely chancy process that determines the selected point without any bias – perhaps a random number generator based on some quantum mechanical event is used to specify a latitude and a longitude. And supposing, as we might, that you believe that this is indeed the selection mechanism, your own subjective probabilities should be correspondingly uniform. Similar points apply to (ii).

You might try to recover the desired answers by taking appropriate limits. You might imagine, for example, a sequence of nested strips around the equator, shrinking to the equator in the limit, and define the conditional probability of W, given E, as the limit of the sequence of ratios of the form $P(W \cap E)/P(E)$. You would need to rule out in a non-question begging way strips that give the wrong answer: for example, we could have converging on the equator a sequence of strips, fat in the west, thin in the east, yielding answers greater than 1/2. Of course, the proposal would not rescue finite cases in which the zero denominator problem arises – e.g., conditioning on the die landing odd. But more to the point, the proposal is *not* the ratio analysis, but rather a substantially more complicated analysis in which ratios appear.

Note that the assumption of *uniformity* in the last examples is not essential to producing the zero denominator problem; it is only there to make the answers obvious. Since there are uncountably many propositions to distribute probabilities over, a real-valued probability measure must assign 0 to uncountably many of them, for it does not assign any infinitesimal values. Note also that it is not the answers themselves that matter, but the fact that *there are answers.* If for some reason you disagree with me on the answers that I have suggested, but agree with me that there are answers, you should agree with me that the ratio analysis is inadequate.

You may be tempted to bite the bullet, saying: "the conditional probability is undefined in these cases – there is no answer, because the ratio analysis gives no answer. If this conflicts with our intuitions, too bad for those intuitions. Our intuitions about cases involving infinities were never that trustworthy, anyway". This is often the right way to respond to a surprising result. Consider the result that the set of even numbers and the set of integers have the same cardinality. This offends the intuitions of many of us. Yet I would not want to run the following superficially parallel argument: "Mathematicians have not adequately analyzed our concept of 'size of set', for they saddle us with an unintuitive result". Rather, I would say that if this result conflicts with our intuitions, too bad for those intuitions. They were never that trustworthy, anyway.

And I might say that of the earth's surface and dart examples, too, if that were the end of the story. But it is not.

4.5. *Kolmogorov's Elaboration of the Ratio Formula*

It turns out that an elaboration of the idea behind the ratio analysis, also due to Kolmogorov, delivers our intuitive answers of 1/2. So our intuitions were apparently quite sound, and in no need of revision. Kolmogorov himself was well aware of the zero denominator problem, and this is what motivated him to elaborate the ratio analysis to handle cases such these. I will briefly describe the Kolmogorov elaboration, though this is not the place to show how it delivers the desired answers.

He introduces the concept of probability conditional on the outcome of a random experiment. The experiment is represented by a sub-σ-field $\mathcal{F}$ of the total σ-field under consideration. We define the random variable $P(A \parallel \mathcal{F})$, *the probability of A conditional on* $\mathcal{F}$, which assumes different values according to which propositions in $\mathcal{F}$ are established as true on the basis of the experiment. $P(A \parallel \mathcal{F})$ is a random variable (whose existence is guaranteed by the Radon-Nikodym theorem) which satisfies the following condition:

$$\text{(RV)} \quad \int_B P(A \parallel \mathcal{F})dP = P(A \cap B) \text{ for all } B \in \mathcal{F}$$

Compare this to the simple transposition of (RATIO):

$$\text{(PRODUCT)} \quad P(A \mid B)P(B) = P(A \cap B).$$

It turns out that this gives us a way of coping with the zero denominator problem, while preserving the guiding idea behind the simpler ratio analysis.[11] We see, then, that we have no need to revise the intuitions that tell us that certain conditional probabilities take particular values in examples such as these.

It might thus be thought that my attack on the ratio analysis is misdirected, and that my real target should be the full-blown Kolmogorov analysis. In response to this, let me firstly observe that it is clearly the ratio analysis that has the limelight in the philosophical literature, and the full-blown analysis is relatively rarely discussed by philosophers.[12] Indeed, various well-known philosophical tomes on probability do not mention Kolmogorov's elaboration (e.g., Fine (1973), Rosenkrantz (1981), Earman (1992), Howson and Urbach (1993), among many others). But more importantly, I think that the full-blown analysis is also inadequate. There are cases in which the probabilities of conjunctions required by the right-hand side of (RV) are vague or undefined, and yet the corresponding conditional probabilities are defined, as we will see in Section 6 and Section 7.

5. INFINITESIMAL PROBABILITIES[13]

You might think that this is a job for non-standard analysis. Recall the regularity theorem: any regular probability function defined on uncountably many mutually disjoint propositions assigns infinitesimal probabilities to uncountably many propositions. Since there are such regular functions, we know that the second horn of the Four Horn theorem is occupied.[14] I will now argue that any function impaled on this horn is in an important sense defective. Infinitesimals do not provide the sanctuary for the ratio analysis that one might hope for.

It will be helpful to begin by returning to our examples of the earth's surface, and the [0, 1] interval, so that we can see how infinitesimal assignments might naturally arise. One might claim that using infinitesimal probabilities in the ratio formula will deliver the correct answers in these examples: "The probability that the chosen point lies on the equator is not 0, but rather some positive infinitesimal, ϵ say; the probability that it lies in the western hemisphere and on the equator is $\epsilon/2$. Likewise, the probability that the dart lands on 1/4 or 3/4 is some positive infinitesimal, twice the probability that it lands on 1/4. (RATIO) then gives the correct answer of 1/2 in each case". One hears this sort of claim a lot in this line of work, and it has considerable prima facie plausibility.

However, I find it wanting, for any putative infinitesimal assignment is *ineffable*. In order to speak of non-standard numbers such as infinitesimals, one must have a non-standard model. We are guaranteed the existence of such a model, if we assume the axiom of choice; but we are also guaranteed that there are *many* of them. We cannot pick out a model uniquely, by naming it or pointing to it. Any talk of '*the* non-standard model' with which we are working is quite misleading. Infinitesimal probability assignments are thus curiously hard to pin down. They are not like familiar probability assignments such as P(the coin lands heads) = 1/2, or even $1/\sqrt{2}$ – there is an ineliminable non-constructive aspect to them. (The appeal to the axiom of choice ought to alert us to this.) I would analogize them more to a gesture of the form 'P(the coin lands heads) = x, for some x in [0, 1]'. This is slightly unfair, perhaps, for we can pin down certain things about them – that they lie strictly between 0 and every positive real number, for instance. Nevertheless, when we are dealing with sets on this tiny scale, where minute differences count for a lot, we want things pinned down *very* precisely. 'ϵ' is a nice name for an infinitesimal probability, but what is it *exactly*? So when philosophers gesture at infinitesimal probability assignments, I want them to give me a *specific example* of one. But this they cannot do; the best they can do is gesture.[15]

For example, one construction of infinitesimals identifies them with equivalence classes of sequences of reals (see Lindstrøm 1988). But what is the equivalence relation? Two sequences are equivalent if they agree term-by-term, 'almost everywhere'. But what does 'almost everywhere' mean? With measure 1. But what is the measure here? It is a finitely additive measure, defined on the power-set of the set $\mathbb{N}$ of natural numbers, that assigns only the values 0 and 1, and that assigns 1 to $\mathbb{N}$ and 0 to all finite sets. But this hardly pins it down! What is the measure of the odd numbers, or the primes, or . . . . ? We cannot fully specify the measure. (In the jargon, this is the problem of the indefinability of an ultrafilter.) But since we cannot fully specify the measure, we cannot fully specify what we mean by 'almost everywhere'; hence we cannot fully specify the equivalence relation; hence we cannot fully specify the equivalence classes; hence we cannot fully specify the infinitesimals. They are thus ineffable.

Note that relative frequencies – even limiting relative frequencies – cannot be infinitesimal. Nor can Popper-style propensities (since they are explicitly defined in terms of such limiting relative frequencies). Those who want to identify objective chances with relative frequencies or such propensities thus cannot appeal to infinitesimal chances to save the ratio analysis in cases such as I have discussed.

In any case, once again, calling infinitesimal assignments "probabilities" at all is a *revision* of the textbook meaning of the word (however they are identified). For as I have said, the notion of countable additivity does not even make sense in this context, so this axiom of Kolmogorov's must be jettisoned. So be it, perhaps – after all, I have no respect for the textbook when it comes to its 'definition' of conditional probability'! But let us then be honest, and explicitly state that the several decades' reign of his axiomatization should come to an end. In any case, I think it is under-appreciated how ill salvaging the ratio formula by means of infinitesimals sits with the rest of the axiomatization.

6. SHARP CONDITIONAL PROBABILITIES, VAGUE UNCONDITIONAL PROBABILITIES

We turn to the next set of trouble spots for the ratio analysis: vague probability assignments.

What is your subjective probability that the Democrats win the next election? If you give a sharp answer, I would ask you to reconsider. Do you really mean to give a value that is precise to infinitely many decimal places? If you're anything like me, your probability is *vague* – perhaps over an interval, but in any case over a range of values. Now, what is your probability that the Democrats win, *given* that the Democrats win? Here there is no vagueness: the answer is clearly *1*. This is our first example of a conditional probability that is sharp, while the unconditional probabilities in the corresponding ratio are vague. Similarly, P(the Democrats do not win, given the Democrats win) = 0; and where **T** and **F** are respectively necessary and impossible, P(**T**, given the Democrats win) = 1, and P(**F**, given the Democrats win) = 0. More interestingly, P(this coin lands heads, given the Democrats win) = 1/2. I will argue that the ratio analysis cannot do justice to examples such as these, yet they could hardly be simpler.

To be sure, we have gone beyond orthodox Bayesianism, according to which opinion is always sharp, always modeled by a single probability function. Let us be unorthodox, then; it is not clear why such sharpness is a requirement of rationality. Your selling price can differ from your buying price for a given bet without rendering you susceptible to a Dutch Book. Your judgments may not determine a unique probability function. And in Jeffrey's decision theory, your probability function may not be fully constrained by your preferences: various utility function/probability function pairs have equal claim to representing you when your utilities are bounded above and/or below. In such cases, there is no single utility function/probability function pair that best rationalizes your behavioral dispositions. So it seems that the attribution of vague opinion may be unavoidable. Moreover, there is a rich body of work on the modeling of such opinion. It might help supporters of the ratio analysis if vague probability can be outlawed, but I put little weight on that reason for doing so.

We have seen how vague unconditional probabilities can be handled "supervaluationally".[16] We consider the set of all of the precisifications of your opinion, your representor. Let D = 'the Democrats win'. Since your probability for D is vague, these precisifications of your opinion assign various different sharp values to D.

What about your sharp conditional probability, $P(D$, given $D) = 1$? If you are a fan of the ratio analysis, you are probably tempted to defend it by saying that for each function p in your representor, the ratio

$$\frac{p(D \cap D)}{p(D)} = \frac{p(D)}{p(D)} = 1.$$

But if you say that, you have only done my work for me: far from defending the ratio analysis, you have displayed that it is inadequate.[17] For while you have used a ratio in your defense, you have used more besides. More generally, the proposal would be to analyze

'your conditional probability for A, given B, is x'

as

'$p(A \mid B) = x$ for every function p in your representor'
(RATIO + REPRESENTOR)

Granted, this is in the *spirit* of the ratio analysis. But it clearly adds something to that analysis, for 'representor' is not a term that appears in the ratio analysis. Notice that I am not (here) saying that (RATIO + REPRESENTOR) is defective; I *am* saying that (RATIO + REPRESENTOR) is not (RATIO), and (RATIO) is my target.

In fact, things are still worse than that. For 'representor' is a term of art that has been introduced to mean the set of all precise probability functions consistent with your state of opinion. But your state of opinion includes both unconditional *and conditional* probability assignments. For instance, if you are like me, you judge that it is a non-negotiable constraint on any conditional probability assignment that the probability of a non-empty proposition, given itself, is 1. Yet this is not reducible to facts about ratios of unconditional probability assignments. (We already saw that with the zero denominator problem, and we will see it again in the next section.) It follows that (RATIO + REPRESENTOR) is circular, since it tacitly makes reference to the very notion of conditional probability that we seek to analyze. The notion of a 'representor' may be useful in modeling vague opinion, but it will not make a good primitive in the analysis of conditional probability.

Note that if chances can be vague, then the ratio analysis of conditional chance is likewise inadequate. For even when the chance of X is vague, various conditional chances, given X, are sharp: the chance of X, given X equals 1, and so on.

Finally, the problem of vague unconditional probabilities alongside sharp conditional probabilities counts equally against Kolmogorov's elaboration of the ratio analysis – for the right hand side of (RV) is vague, yet it is equated with something sharp.

7. DEFINED CONDITIONAL PROBABILITY, UNDEFINED UNCONDITIONAL PROBABILITIES

According to the ratio analysis of 'the probability of A given B', if either or both of the numerator $P(A \cap B)$ and the denominator $P(B)$ are undefined, the conditional probability is undefined. Yet there are many cases in which this is implausible. This unwelcome result for the ratio analysis is avoided only by incurring undesirable commitments.

It is this argument that I want to give the most attention. In this section, I will offer two examples that I will discuss at some length. The first will be a quite straightforward example involving coin tossing. The second will be more technical, involving as it does non-measurable sets. It has the virtue of mathematical rigor, and it shows that conditional probabilities can be well-defined even when the corresponding terms of the ratio *cannot* be defined, consistent with certain natural constraints. In the next section, I will discuss a series of instances drawn from actual scientific and philosophical practice in which probability gaps arise. In both sections, I will contend that my argument goes through for subjective and objective probability alike.[18] We will thus see yet again that the relevant horn of the Four Horn theorem is occupied by many interesting probability functions.

7.1. *Subjective Probability*

In this section, we will interpret 'P' to be the credence function of a rational agent: 'you'.

7.1.1. *A Coin Example*

Here is a coin that you believe to be fair. What is the probability that it lands heads ('Heads'), given that I toss it fairly ('Toss')? 1/2, of course. According to the ratio analysis, it is $P(\text{Heads} \mid \text{Toss})$, that is:

$$\frac{P(\text{the coin lands heads} \cap \text{I toss the coin fairly})}{P(\text{I toss the coin fairly})}.$$

However, I was careful not to give you any information on which to base these unconditional probabilities. Such information is no part of the specification of the problem. Take P(I toss the coin fairly). Perhaps I hate tossing coins, and would never toss one; perhaps I love tossing coins, but hate tossing them fairly; perhaps I haven't made up my mind whether I will toss it fairly or not, but I will spontaneously decide one way or another a minute from now; perhaps I will throw a dart at a representation of the [0, 1] interval, and toss it fairly or not on the basis of where it lands. But I have told you nothing about that. I believe that it is quite rational for you, then, not to assign any value to this probability whatsoever. The same applies to P(the coin lands heads $\cap$ I toss the coin fairly). The terms in the ratio, then, remain undefined; and 'undefined' divided by 'undefined' does not equal 1/2.

The example could hardly be simpler, yet the ratio analysis falters on it. However, I want to address three objections that one might have at this point.

Objection i. "The unconditional probabilities are in fact defined"
It might be objected that $P(\text{Heads} \cap \text{Toss})$ and $P(\text{Toss})$ are defined after all. Let us begin with the weakest version of this objection. One might say that, given your lack of information, you would assign $P(\text{Toss}) = 1/2$, and $P(\text{Heads} \cap \text{Toss}) = 1/4$, by the symmetry of the situation. For example, you assign the value to $P(\text{Toss})$ as follows. Either I toss the coin or not, and there is no basis for favoring one hypothesis over the other, so by the principle of indifference, each hypothesis should get the same probability, namely 1/2. And given that the coin is fair, it follows that $P(\text{Heads} \cap \text{Toss}) = 1/4$.

The principle of indifference famously yields inconsistent results. To see that here, consider now the hypothesis that I toss the coin while whistling "Waltzing Matilda". Either I do or I don't, and you have been given no information either way, so by the principle of indifference, each hypothesis should get the same probability, namely 1/2. Now consider the hypothesis that I toss the coin while singing "The Star Spangled Banner". By the principle of indifference, you should assign probability 1/2 to that as well. I would not recommend it.

One might insist that you must assign a probability to 'Toss' nonetheless – perhaps based on your prior knowledge about people's proclivity towards fair coin tossing, or what have you. I still find this absurd. For starters, I would ask you to be honest: you *didn't* assign any value to 'Toss', did you? But even if you did, for some reason, why *must* you assign a value to 'Toss'? No one is coercing you to make bets on 'Toss', or to form corresponding preferences; no one is coercing you to make the relevant judgments; no one is coercing you to form the relevant dispositions. And if someone did coerce you, we would get an answer alright, but it is doubtful if it would reveal anything about your state of mind prior to the coercion. In short: why not leave 'Toss' as a probability gap? In the words of Fine (1973, p. 177), "I think it wiser to avoid the use of a probability model when we do not have the necessary data than to fill in the gaps arbitrarily; arbitrary assumptions yield arbitrary conclusions".

Moreover, you might have principled reasons for *refusing* to fill in the gaps for 'Toss' and for 'Heads ∩ Toss'. They both apparently involve the performance of a free act. Now it is not clear that free acts are the sorts of things to which objective probability attaches. That is, they may well be objective probability gaps – more on this in Section 7.2.2. And if you believe them to be objective probability gaps, then you might well deliberately refrain from assigning them subjective probability (except under coercion . . .).

Or following Spohn (1977), Kyburg (1988), Gilboa (1994), and Levi (1997) among others, perhaps you think that there is no obstacle to assigning probability to the free choices of other people, but that you cannot assign probability to your *own* free choices – more on this in Section 8.4. Then let us trivially modify the example so that it is *you*, rather than I, who chooses whether or not to toss the coin. The gap returns.

Or without even getting into possibly thorny metaphysical issues about chance, or about free agency, change the example again: What is the probability that it lands heads, given that Joe tosses it fairly? 1/2, of course. But there is no fact of the matter of the probability that Joe tosses the coin fairly. *Who is this Joe, anyway?* None of that matters, however, to the conditional probability, which is well-defined (and obvious). By analogy, we can determine that the argument:

> Joe is a liar
> Therefore,
> Joe is a liar

is valid, even though there is no fact of the matter of the truth value of the statement 'Joe is a liar'.

The next objection can be dispensed with more quickly.

Objection ii. "The ratio can be defined without the terms in it being defined"

You might grant me that the numerator and denominator of the ratio $P(\text{Heads} \mid \text{Toss})$ are not defined, and hence not known – but maintain that nonetheless the ratio is known to be 1/2, and hence is defined. You might say: "It is not uncommon for one to know the value of a ratio without knowing the values of the terms that enter into it. For example, one can see two children perfectly balanced on the ends of a seesaw (teeter-totter), and know that the ratio of their weights is 1, without having a clue about their individual weights. So the ratio analysis can deliver the correct answer here, despite the absence of information about the unconditional probabilities that enter into it".[19]

The seesaw analogy is inapt. Even though we do not know the respective weights of the children, we know that *they do have weights*. This is just as well, both for the sake of the children, and for the sake of the ratio of their weights. For it would be nonsense to say "the ratio of the children's weights is 1, but the children don't have weights". Rather, one says "the ratio of the children's weights is 1, but I don't know what those weights are". On the other hand, it is not the case that there are true but unknown values of your subjective probabilities for 'Heads ∩ Toss' and for 'Toss'.

That makes it sound like these values inhabit some shadowy Freudian realm of your subconscious. If that turns out to be the case, it is hardly something that we can derive simply by reflection on the ratio formula! And how could anyone be so confident that these putative values stand in the right ratio if they remain unknown? Anyway, the putative unconditional probabilities that the ratio analysis forces upon us simply may not exist, as I have argued. And it is nonsense to say "the ratio of the unconditional probabilities is 1/2, but the unconditional probabilities do not exist".

Objection iii. "Supervaluate"

This reply will be familiar. It grants me that there are gaps for the unconditional probabilities of 'Heads ∩ Toss' and of 'Toss', but insists that every precisification of these values yields a ratio of 1/2: "While there is discord among the functions p in your representor regarding the values of p(Heads ∩ Toss) and p(Toss), they speak with one voice regarding the value of p(Heads | Toss), saying '1/2' in unison".

And my counter-reply will be familiar. This reply gives up on the ratio analysis of the conditional probability here, for it proposes that we *replace* that analysis with

$p(\text{Heads} \mid \text{Toss}) = 1/2$ for every function p in your representor'.

So this isn't really an *objection* to me at all. Moreover, as I argued in Section 6, the term 'representor' requires the notion of conditional probability in its analysis, rather than the other way round.

The 'supervaluational' approach essentially tries to identify *absence of probability* with *presence of vague probability*. They are not the same thing. I have argued that you may rationally assign no probability whatsoever to 'Toss'. Any precisification then spells out ways that your probabilities *determinately are not.* Any filling of your gap for 'Toss' is a distortion of your true state of mind – speaking, where you choose to remain silent.

7.1.2. *Non-Measurable Sets*

This brings us to another case in which silence is golden. Recall my discussion in Section 2.3 of the uniform probability measure over [0, 1], and its associated non-measurable sets. I gave a proof of the existence of non-measurable sets – in fact, denumerably many of them: the choice set C, and all of the translations C_r of it by a rational distance r. Now, the proof did assume countable additivity, and you may balk right there. However, as I have said, that is tantamount to balking at Kolmogorov's axiomatization of probability. So let me proceed for now under the assumption that you are less radical than that.

Imagine again throwing an infinitely fine dart at the [0, 1] interval, with you assigning a uniform distribution (Lebesgue measure) over the points at which it could land. What probability do you give to its landing in C? Surely the answer is: there is no answer. C is the archetypal probability gap; $P(C)$ is undefined.

Now, what is the probability of the dart's landing in C, given that it lands in C? As usual, the answer is 1. So we have $P(C$, given $C) = 1$. What we do *not* have is $P(C \mid C) = 1$. Rather,

$$C \cap C = C,$$

so

$$P(C \mid C) = \frac{P(C)}{P(C)} = \frac{\text{undefined}}{\text{undefined}}.$$

And undefined/undefined is not 1. (You should resist any temptation to say: "the ratio is 1, since everything is self-identical, and so $P(C) = P(C)$" – rather, everything *that exists* is self-identical, but the probability of C does not exist.)

We know how to multiply the examples. Here's another one: $P(\text{not-}C$, given $C) = 0$. This case is a little different, because now the ratio analysis would have us compute a ratio in which only one of the terms is undefined:

$$P(\text{not-}C \mid C) = \frac{P(\text{not-}C \cap C)}{P(C)} = \frac{0}{\text{undefined}} = \text{undefined}.$$

As before, let **T** be necessary: $P(\mathbf{T}$, given $C) = 1$, while $P(\mathbf{T} \mid C)$ is undefined. Let **F** be impossible: $P(\mathbf{F}$, given $C) = 0$, while $P(\mathbf{F} \mid C)$ is undefined. And as before we can give non-degenerate examples: P(a fair coin lands heads, given C) = 1/2, while P(a fair coin lands heads $\mid C$) is undefined. In fact, it seems that for every possible value, we can produce a conditional probability that intuitively has that value, yet that has no value according to the ratio analysis. Suppose a second dart is also thrown at [0, 1], independently of the first. For each real number r in the interval, we have

$$P(\text{the second dart lands inside } [0, r], \text{ given } C) = r.$$

Again and again (uncountably many times!), the ratio analysis fails to yield the correct answer – or, indeed, any answer.

You might say that since it fails to yield an answer, there is no answer – that conditional probabilities go undefined when the conditions are non-measurable. This is another 'bullet'-biting response, reminiscent of the one I discussed in connection with the zero denominator problem. But again it strikes me as quite unsatisfactory. When I held it as self-evident that the conditional probability of every (non-empty) proposition, given itself, is 1, no proposition (apart from the empty one) was exempt. Every proposition implies itself, however strange that proposition might be in other respects. C is no exception. The truth of C, *conditional on* the truth of C, is guaranteed.

'Supervaluating', I have argued, revises rather than rescues the ratio analysis; but in any case it will not make the problem go away. We could try to save the ratio analysis by considering each possible assignment of a precise value to C, and claiming that the ratios above come out right on each such assignment. (Well, not quite *each* such assignment, since we can't allow an assignment of 0 to C – the zero denominator problem strikes again!) We could, that is, consider extensions of our measure that assign a value to C, where previously it received none. But we cannot replicate the given probability assignment for each translated counterpart C_r of C – that was the point of the proof of the existence of non-measurable sets, after all. Yet translation invariance was an inviolable constraint on the distribution. In short, each 'valuation' would distort the original probability distribution, making it 'lumpy' rather than uniform.

The point is this: *it is impossible to meet both the unconditional and the conditional probability (as analyzed by the ratio) constraints at the same time.* If we respect the unconditional probability assignments, the conditional probability assignments are not respected: where we should be getting well-defined answers such as 1 and 0, we get no answer at all. If we respect the conditional probability assignments (as analyzed by the ratio), the unconditional probability assignments are not respected: a distribution that is supposed to be uniform turns out to be lumpy. A supporter of the ratio analysis is then forced to say that there is no such thing as a uniform (i.e., translation invariant) distribution over [0, 1]. But that is a strange thing to say.

Incidentally, not even Kolmogorov's elaboration of the ratio analysis will help here. The crucial formula, remember, was this:

$$\text{(RV)} \qquad \int_B P(A \parallel \mathcal{F})dP = P(A \cap B) \text{ for all } B \in \mathcal{F}.$$

We cannot substitute C for both A and B, because C does not belong to the relevant sigma-algebra (being non-measurable). There is no sense to be made of the right-hand-side of this equation in our example, because $P(C \cap C)$ has no value.

There are ways to avoid non-measurable sets, but they are not cost-free. For example, if we drop countable additivity and assume only finite additivity, we can restore measurability of all sets in the power set of our sample space. But finite additivity leads to other headaches, notably failures of conglomerability (Schervish et al. 1984). So non-measurable sets won't go away so easily.

One might concede all of this, but regard these examples as 'don't cares'. One might call them strange, or even pathological. But calling them names won't make them go away; and if we want to develop a rigorous and fully general probability theory, then we are stuck with them. In fact, we are stuck with *many* of them – for it can be shown that every set with positive Lebesgue measure contains a non-measurable set.

More to the point, I do not accept that they are so strange. For a set that is non-measurable with respect to one measure is measurable with respect to another.[20] In this sense, 'non-measur*able*' is quite a misnomer. Calling a set 'non-measurable' makes it sound as if non-measurability is both an intrinsic and a modal property of the set, one that it has once and for all; this in turn invites the name-calling. I want to stress that non-measurability is a *relation* that a set may bear to one probability measure, while not bearing it to another.

To be sure, one measure might be especially salient, as the Lebesgue measure is in our dart example. Still, even fixing that measure, *the dart must land in some non-measurable set* – wherever it lands, it lands inside C_r for *some* r. After all, the $\{C_r\}$ form a partition of [0, 1]. So in this sense, there is simply no avoiding 'strange' things happening. But this fact is utterly commonplace! It is as mundane as the fact that your body currently occupies a certain salient region of space B, while also lying inside various gerrymandered regions of space, such as the mereological sum of B, the Eiffel Tower, and Alpha Centauri. And far from being of marginal interest, the existence of non-measurable sets is one of the most important results of measure theory. Non-measurable sets also have interesting applications, as we will see shortly.

7.2. *Objective Probability*

That completes the problem of undefined unconditional probabilities, defined conditional probabilities, in the case of subjective probability. Now I want to show that the problem remains in the case of objective probability. Again, we see that the ratio analysis fails for conditional chance.

In this subsection, I will do things in the opposite order from the last subsection. That is, I will start with a brief reprise of the non-measurable sets problem, since that is fresh in our minds, then revisit the coin problem, then proceed to further problems.

7.2.1. *Non-Measurable Sets, Revisited*

I can recast the problem of non-measurable sets very quickly. After all, the proof of the existence of non-measurable sets was neutral as to the interpretation of probability. So all of the things I said before now get construed in terms of objective probabilities. We can think of the dart-throwing as an objectively chancy process, with a uniform chance distribution over the [0, 1] interval. Since chance supposedly conforms to the probability calculus, we simply repeat the argument that showed the existence of non-measurable sets, such as C. Now, C is understood as a *chance gap*. The conditional chance of the dart landing in C, given it lands in C, is 1; but this cannot be represented as the ratio of the chance of C to itself, for there is no such thing as the chance of C. Likewise for the other examples that I gave in the previous sub-section.

What's more, objective conditional probabilities with non-measurable conditions have proved to be of significant use. Witness Pitowski's (1989, Ch. 5) employment of them in modeling quantum statistics with "realistic" local hidden variable theories, according to which the violations of classical Bell-type constraints occur because of the non-measurability of certain events (associated with subsets of [0, 1]).

7.2.2. *The Coin Example, Revisited*

Now let's return to the coin example. It is reasonable to suppose that there is a certain conditional chance that the coin lands heads, given that I toss it fairly – indeed, supposing the coin to be fair, that chance is surely 1/2, at least at any time t up to the moment the toss begins. The advocate of the ratio analysis equates the conditional chance with the ratio of two unconditional chances, $P_t(\text{Heads} \cap \text{Toss})$ and $P_t(\text{Toss})$, for all such t. But it is not clear that these chances are all defined, for they involve propositions about free acts.

The conviction that chances attach to propositions about free acts, such as 'Toss', is perhaps best justified by appealing to the grander conviction that *every* proposition has a chance attached to it, and then finding a justification for *that*. After all, if there are counterexamples to the grander conviction, then propositions about free acts may well be good candidates to be among them.

Indeed, the advocate of the ratio analysis *had better* justify the grander conviction – otherwise I will simply appeal to the existence of some chance gap G, whatever it may be, and then use $P(G, \text{given } G)$ to make my point. But showing that there are no chance gaps will be no easy task, since this is a significant and contentious metaphysical contention. Why think it? Levi (1980) doesn't; van Fraassen (1989) doesn't. Cartwright (1999) doesn't, and indeed she thinks that chances only arise in rather special circumstances so that, in my terminology, chance gaps predominate. I am agnostic about their ubiquity – partly because I am not sure what chance exactly *is* in the first place, even though I believe that there is such a thing. But I think that it is unfortunate for the advocate of the ratio analysis who also believes there is such a thing, that her advocacy forces her into faith in that contention, even though she may not be sure what chance exactly is either. The advocate may say that her prior belief in the truth of the ratio analysis justifies, and indeed implies the truth of the claim that all propositions have chance values. I find it surprising that seemingly innocent statements about conditional chance should foist upon us a commitment to a controversial view about chance. Maybe that view is right; but it is surely not implied by our prosaic attributions of conditional probability. It is striking that such an interesting metaphysical conclusion can be derived from such a humble starting point.

If chance is left unanalyzed, then we are left totally in the dark as to why it should attach to every proposition. If it is analyzed as a 'propensity' or a 'disposition', then this only redescribes the problem: why should we believe that every proposition has a propensity/disposition? And if it is analyzed as relative frequency in the infinite long run, then it seems that we can make sure that it *doesn't* attach to every proposition. To see this, we can exploit the freedom of the act of tossing the coin. Suppose that each day I can either (fairly) toss the coin or not, and imagine a hypothetical infinite sequence of such trials, each of which involves my free choice. Now suppose that I deliberately make a sequence of choices so that the successive relative frequencies of 'I toss the coin' days, over the total number of days, fluctuates wildly. Here is one such sequence, in which I toss the coin, then do not toss the coin, for runs of exactly 2^n consecutive days (abbreviating 'Toss' by 'T'):

$$T\bar{T}TT\bar{T}\bar{T}TTTT\bar{T}\bar{T}\bar{T}\bar{T}TTTTTTTT\bar{T}\bar{T}\bar{T}\bar{T}\bar{T}\bar{T}\bar{T}\bar{T}\ldots.$$

The relative frequency of T-days does not approach a limit, so by a relative frequency analysis of chance, the probability of T is not defined. And yet corresponding conditional probabilities can be defined, as (limits of) relative frequencies. For example, the probability of T, given T, is 1: the (limiting) relative frequency of T-days among T-days is 1.

We have here in turn an argument for thinking that at least some free acts could lack chance values. For as I said earlier, while we should not identify chance with relative frequency, still the latter is typically a good guide to the former. The fact that we can in principle guarantee that our free choices fail to yield a convergent relative frequency suggests to me that these choices *could* be chance gaps. Admittedly, there are difficulties here in defining a kind of trial, and perhaps we might want to say instead that it is the entire sequence that gets assigned a chance, namely 1; and there are thorny issues in the compatibilist/incompatibilist debate that I cannot get into here. But to repeat, it is the friend of the ratio, and not I, who must take a stand on the matter.

There is always, of course, the bullet-biting reply: we have no assurance that the probability of Heads, given Toss, is defined after all. More generally, the reply has it that all chances conditional on free acts are undefined if the acts themselves lack chance values. But as we will see in the next section, yet again the bullet is not so easily bitten.

8. EXAMPLES FROM SCIENCE AND PHILOSOPHY

I have given what I take to be clearly the right values to the conditional probabilities in these various examples, and the many people whom I have asked almost invariably give the same answers. I take this to be indicative of some broadly shared notion, derived from commonsense, though refined by philosophical reflection; and while such notions are defeasible to be sure, they should not be given up lightly. But I do not want to rest my case on intuitions, widely shared or otherwise, refined or otherwise. So let us now turn to various examples drawn from actual scientific and philosophical practice in which conditional probabilities are defined even when the corresponding unconditional probabilities may not be. The first example concerns putatively objective probabilities; the second concerns both objective and subjective probabilities; the rest concern subjective probabilities. They are far from exhaustive.

8.1. *Quantum Mechanical Probabilities*

Quantum mechanics apparently tells us that certain chances, conditional on free acts, are defined, and it even purports to tell us their values. For example, it tells us that the chance that a certain particle is measured to be spin-up, given that it is measured for spin in a given direction, is 1/2; and the condition surely involves a free choice. Recall the response that I have called "biting the bullet" when (RATIO) goes undefined. Here it would be that quantum mechanics may actually be *incorrect* in its attribution of well-defined conditional probabilities to the results of certain measurements, because these probabilities may in fact be undefined. But, to understate the point, if these probabilities are undefined, one wonders why the statistics from the laboratory confirm the quantum mechanical probabilities so well, or indeed why those statistics should have any stability at all. Needless to say, I am underimpressed by such armchair suspicion of one of our best scientific theories.

Let me elaborate. A species of conditional probability makes its way into quantum mechanics via the Born rule. In its simplest form, this rule has three essential elements: a proposition M about some measurement taking place on a system (in a certain state); a proposition O_k about which outcome of this measurement eventuates; and a probability assignment p_k to this outcome proposition, conditional on the measurement. If this conditional probability is not to be analyzed by the ratio formula, then my point is already made. But it is commonly thought that it *should* be so analyzed:[21]

$$P(O_k \mid M) = \frac{P(O_k \cap M)}{P(M)} = p_k.$$

This conditional probability is standardly regarded as being objective. Again, for this to be correct, the probabilities that appear in the numerator and denominator must be well-defined.

Quantum mechanics itself does not provide these probabilities.[22] For example, it does not give probabilities to propositions of the form 'measurement X is performed by an experimenter'.[23] Perhaps you think otherwise: "Quantum mechanics *does* ultimately provide probabilities for such measurements being performed – for an experimenter is just a large collection of atoms, and is thus a (hugely complicated) quantum system. Likewise the measurement apparatus; likewise the combined system of experimenter plus measurement apparatus ..." But I contend that quantum mechanics *cannot* deliver even in principle probabilities of the form $P(M)$. I have two reasons for saying this.

Firstly, quantum mechanics only delivers further conditional probabilities, according to Born's rule, of the very form that the ratio was supposed to analyze. It does not assign values to *any* unconditional probabilities of the form $P(X)$, irrespective of what X is; *a fortiori* it does not assign values to unconditional probabilities of the special form $P(M)$.

Secondly, let us not forget just how special the M propositions are. They are propositions about *measurement processes*, and so assigning a probability to them may be especially problematic. While Born's rule assigns (conditional) probabilities to measurement outcomes, it surely does not assign probabilities (conditional or unconditional) to the processes that lead to those outcomes. Indeed, the latter would seem to require the assignment of probabilities by Born's rule to the decisions of physicists as to whether or not the relevant measurements are performed! Is any such decision itself the outcome of some further 'super-measurement', as standard interpretations of quantum mechanics would seem to require? This not only cries out for an account of super-measurement (for it is *prima facie* implausible that there are such processes, or at least unclear what they would be), but it also adds bite to the measurement problem. After all, until this super-measurement is made, the physicist should be in a superposition of two states, namely 'decides to perform' and 'decides not to perform'. This hardly accords with the facts as we take them to be. In any case, how are we to analyze these putative conditional probabilities of the form "the probability that measurement X takes place, given super-measurement S takes place"? The ratio analysis would commit us to unconditional probabilities of the form P(super-measurement S takes place). Infinite regress threatens if we try to derive these in turn from Born's rule – for it seems we must now invoke 'super-super-measurements', of which super-measurements are the outcomes, and so on.

The advocate of the ratio analysis may concede these points, but may insist nonetheless that these unconditional probabilities really do exist, and yield the correct ratios, even though quantum mechanics itself does not provide them. For example, most of the main solutions to the measurement problem (Bohm's interpretation, collapse interpretations, the many worlds interpretation, and so on) have the consequence that probabilities for measurements taking place are well-defined. I say "most" – instrumentalists about quantum mechanics, who say roughly that it is just a calculational device, and that it is not intended to provide a complete description of the physical, can agree that it yields conditional probabilities without there necessarily being any probability for the condition M.[24] Still, the advocate insists, if any of the other main solutions is correct, this problem for the ratio analysis is solved.

I have two replies to this. The first you have already seen. M is typically a proposition about the (apparently) free act of an experimenter – whether or not he or she orients a Stern-Gerlach apparatus a certain way, and prepares a particle in a suitable state, and so on. As such, M may not have a chance value at all. Or it may. But whether it does or not is controversial, not agreed upon on by all interpretations of quantum mechanics, and certainly not something that the ratio formula by itself can settle. It seems to me that the intuition that chances must always exist, even for free acts, parallels the intuition that values for observables (such as position and momentum) must always exist. But the latter intuition has been challenged since Bohr, and has hit particularly hard times since the Kochen-Specker theorem.

Secondly, if $P(M)$ has some determinate value, then this a macroscopic fact that is seemingly not reducible to the facts that quantum mechanics describes. If that theory is a complete description of the microscopic realm, then it seems that we have macroscopic facts that are not reducible to microscopic facts. So it seems that the ratio analysis commits us either to the incompleteness of quantum mechanics as a theory of the microscopic world (move over, Einstein, Podolsky and Rosen!), or to the existence of macroscopic facts that cannot be reduced to microscopic facts. Either way, we find an *empirical* consequence of the ratio analysis of Born's rule that we may prefer not to have foisted upon us. Quantum mechanics is thought by many to be a complete theory of the micro-physical; and it is natural to think of macroscopic facts as large conjunctions of microscopic facts. At any rate, if these positions are to be ruled out, that is the job of nature, not of conceptual analysis, let alone conceptual analysis of the concept of conditional probability.

8.2. *Probabilistic Causation*

There is something appealing to the idea that causes raise the probabilities of their effects. And it is appealing to express this idea in terms of inequalities among conditional probabilities. The simple-minded formulation

$$A \text{ causes } B \text{ iff } P(B, \text{given } A) > P(B, \text{given not-}A)$$

cannot be quite right (witness spurious correlations and the asymmetry of the causal relation), even though it is already a substantial step in the right direction. We may go on to refine it in various ways, for example:

$$A \text{ causes } B \text{ iff } P(B, \text{given } A \cap X) > P(B, \text{given not-}A \cap X)$$
$$\text{for every member } X \text{ of some 'suitable' partition} \ldots$$

The exact details are not important for my point. What *is* important is that many authors have used the ratio formula for such conditional probabilities (see e.g., Suppes 1970; Cartwright 1979; Salmon 1980; Eells 1991). This, in turn, commits them to the existence and sharpness of the corresponding unconditional probabilities – an implausible commitment. You and I know that smoking causes lung cancer – in some appropriately formulated sense, the probability that I get lung cancer, given that I smoke, is greater than the probability of lung cancer, given that I do not smoke. We know it, even though we may assign no probability to the proposition *that* I smoke. Indeed, we could make similar causal claims about our coin-tossing friend *Joe*.

8.3. *The Semantics of Indicative Conditionals: Adams' Thesis*

Huw Price (1986) notes the wide acceptance of Adams' Thesis that the assertability of an indicative conditional goes by the subjective conditional probability of the consequent, given the antecedent. For example, I regard as highly assertable the conditional 'if it is raining in Moscow, the ground in Moscow is wet' because I assign high conditional credence to the ground in Moscow being wet, given that it is raining in Moscow. Price goes on to argue that the truth of Adams' Thesis is inconsistent with (RATIO):

> For it entails that a person cannot properly be in a position to assent to an indicative conditional unless they hold some degree of confidence in the antecedent and in the conjunction of the antecedent and the consequent (leaving aside the requirement that these two degrees stand in an appropriate ratio). Whereas in fact, of course, we often feel perfectly justified in assenting to such a conditional when we simply have no opinion as to whether the antecedent is true ... I think it is clear that forming a considered judgement about an indicative conditional need not involve forming such a judgement about its antecedent. The conditional answers to different standards; which means that if assent to the conditional goes by conditional credence, conditional credence does not go by [RATIO]. (21)

The subjective conditional credence of the consequent C, given the antecedent A, that figures in Adams' Thesis is invariably represented by the ratio $P(C \mid A)$. As Price observes, this requires the corresponding unconditional credences to be defined. Yet we frequently can judge the assertability of an indicative conditional in complete absence of such credences – I can regard the conditional concerning Moscow as assertable without assigning a degree of belief to 'it is raining in Moscow'. It follows, then, either that this assertability does not go by the conditional credence after all, or that the conditional credence does not goes by the ratio. But Adams' Thesis is thought to *sound right*, and to accord with our intuitions about a wide range of cases. What sounds right and so accords, I suggest, is equating the assertability of the conditional to the corresponding conditional credence; it is the further equating of this conditional credence with a ratio of credences, then, that is the culprit.

8.4. *Decision Theory and Game Theory*

Decision theory tells you how your beliefs and desires in tandem determine what you should do. It combines your utility function U and your probability function P to give a figure of merit for each possible action, called the *expectation*, or *desirability* of that action. Let the possible outcomes be $\{O_j\}$, and your possible actions $\{A_i\}$. Evidential decision theory, as presented by Jeffrey (1983b), explicitly uses the ratio analysis in its expectation formula:

$$\text{Desirability}(A_i) = \sum_j U(O_j \cap A_i).P(O_j \mid A_i).$$

But it is not part of that theory that you assign probabilities *to the actions A_i themselves*. Indeed, certain authors are quite explicit that you *need not*, and even *cannot* assign them probabilities. Price puts it this way: "It is absurd to suppose that in deciding whether to act in a certain way, an agent makes (or should make) use of a degree of confidence that he or she will so act" (22). Edgington (1995) echoes these remarks:

> ... we often have a degree of belief in B given A when we have not determined what we think about A. One important case is when I am deliberating about what to do: A has the form "I do x", and B is a possible consequence of doing x. It would be absurd to hold that I have to figure out how likely it is that I will do x, before I can arrive at a judgement [$P(B$, given $A)$]. (266)

Mellor (1993, p. 236) makes much the same point. Kyburg (1988, p. 80) goes further, contending that "to the extent that I am actually making a choice, I must regard that choice as free ..." In doing so, he continues, I *cannot* assign probabilities to my acting in one way rather than another (even though onlookers may be able to do so). Spohn (1977), Gilboa (1994), and Levi (1997) all argue for variants of this position. It is unclear, for example, how one can make sense of non-trivial subjective probabilities of one's future actions on the betting interpretation: one's fair price for a bet on such an action ought to collapse to 0 or 1. If these authors are right, then the conditional probabilities that appear in the expectation formula are not to be evaluated by the ratio formula.

Similar points carry across to game theory, in which outcomes are determined partly by the actions of other players. While you have subjective probabilities over the actions of all the *other* players in a given game (according to the subjectivist interpretation of the Nash equilibrium), you need not, and perhaps cannot have probabilities over your *own* actions. Nonetheless, conditional probabilities for the various outcomes, given your own actions, are still well-defined.

8.5. *Expert Probabilities*

We are often guided by the opinions of experts. We consult our doctors on medical matters, our weather forecasters on meteorological matters, and so on. Gaifman (1988) coins the terms "expert assignment" and "expert probability" for a probability assignment that a given agent strives to track:

"The mere knowledge of the [expert] assignment will make the agent adopt it as his subjective probability" (193). This idea may be codified as follows:

(Expert) $P(A \mid pr(A) = x) = x$, for all x such that $P(pr(A) = x) > 0$,

where 'P' is the agent's subjective probability function, and '$pr(A)$' is the assignment that the agent regards as expert. For example, if you regard the local weather forecaster as an expert on your local weather, and he assigns probability 0.1 to it raining tomorrow, then you may well follow suit:

$$P(\text{rain} \mid pr(\text{rain}) = 0.1) = 0.1.$$

More generally, we might speak of an entire probability function as being such a guide for an agent over a specified set of propositions. Van Fraassen (1989, p. 198) gives us this definition: "If P is my personal probability function, then q is an *expert function for me concerning* family F of propositions exactly if $P(A \mid q(A) = x) = x$ for all propositions A in family F".

Let us define a *universal expert function for* a given rational agent as one that would guide *all* of that agent's probability assignments in this way: an expert function for the agent concerning all propositions. Van Fraassen (1984, 1995), following Goldstein (1983), argues that an agent's *future probability functions* are universal expert functions for that agent. He enshrines this idea in his Reflection Principle, where P_t is the agent's probability function at time t, and $P_{t+\Delta}$ is her function at a later time $t+\Delta$:

$P_t(A \mid P_{t+\Delta}(A) = x) = x$, for all A and for all x such that $P_t(P_{t+\Delta}(A) = x) > 0$.

The principle reflects a certain demand for diachronic coherence imposed by rationality. Van Fraassen (1984, 1995a) defends it with a diachronic Dutch Book argument, and by analogizing violations of it to the sort of pragmatic inconsistency that one finds in Moore's paradox.

We may go still further. There may be universal expert functions for *all* rational agents. Let us call such a function a *universal expert function*, without any relativization to an agent. The *Principle of Direct Probability* regards the *relative frequency* function as a universal expert function. Let A be an event-type, and let *relfreq*(A) be the relative frequency of A. Then for any rational agent with probability function P, we have

$P(A \mid relfreq(A) = x) = x$, for all A and for all x such that $P(relfreq(A) = x) > 0$.

(Cf. Hacking 1965.)

Lewis (1980) posits a similar universal expert role for the *objective chance function*, *ch*, in his *Principal Principle*:

> $P(A \mid ch(A) = x) = x$, for all A and for all x such that $P(ch(A) = x) > 0$.[25]

A frequentist who thinks that chances just *are* relative frequencies would presumably think that the Principal Principle just *is* the Principle of Direct Probability; but Lewis' principle may well appeal to those who have a very different view about chances – e.g., propensity theorists.

I have so far followed these authors in presenting these constraints in terms of the ratio analysis. But what is intended, of course, are constraints on conditional probabilities of the form $P(A,\ \textit{given}\ q(A) = x) = x$. Consider again the Reflection Principle. This constraint presumably obtains whether or not you assign sharp unconditional probabilities to propositions about the values of $P_{t+\Delta}(A)$ for every A. Dutch bookies cannot fleece you for failing to assign such probabilities; there is no pragmatic inconsistency in failing to assign such probabilities. The Reflection Principle, then, may well be a constraint on conditional probability assignments, but I doubt that it is a constraint on ratios of unconditional probability assignments.

Or consider again the Principal Principle. It is not at all clear that rationality demands that, for every A and for every x, a certain ratio of unconditional probabilities take a particular value:

$$\frac{P(A \cap ch(A) = x)}{P(ch(A) = x)} = x.$$

It isn't just that written this way, as a rather complicated ratio of probabilities, the Principal Principle loses its immediate plausibility (though that too is true). The principle is presumably a constraint on rational opinion, whether or not you assign sharp unconditional probabilities to propositions about the values of $ch(A)$. Moreover, I doubt that it is a constraint on rational opinion *that* you assign such probabilities.

And so it goes. The ratio analysis misrepresents the commitments we undertake when we regard various probability assignments, or functions, as experts.

9. THE RATIO: BAD ANALYSIS, GOOD CONSTRAINT

Let us pause for a moment to take stock. My strategy has been to try to push proponents of the ratio analysis into corners that they might rather not occupy. Along the way, I have had my various imaginary – and in some cases, real-life – respondents find that their commitment to the ratio analysis has also committed them to saying such things as:

- "the chance function (at any given time) is regular, or conditional chances of the form $P(X$, given $Z)$ are never defined when $P(Z) = 0$";
- "all rational credence functions are regular, or for all rational credence functions P, conditional probabilities of the form $P(X$, given $Z)$ are never defined when $P(Z) = 0$";
- "Kolmogorov's axiomatization of unconditional probability should be rejected";
- "total ignorance should be modeled with sharp probability rather than with probability gaps";
- "vague probability should be outlawed";
- "there is no such thing as a uniform distribution over [0, 1]";
- "there are no chance value gaps";
- "either quantum mechanics is incomplete as a theory of the microscopic world, or there exist macroscopic facts that cannot be reduced to microscopic facts";
- "rationality requires one to assign probabilities to:
  - all events that figure in causal relations that one recognizes;
  - the antecedents of all the indicative conditionals whose assertability we can judge;
  - all of one's own future choices among which one is deciding;
  - all propositions about one's expert assignments: all propositions about one's own future probability assignments; all propositions about the values of the chance function at any given time; and so on".

This completes my critique of the ratio as an analysis of conditional probability. I have identified four varieties of trouble spots for the ratio analysis: zero probability assignments, infinitesimal assignments, vague probability assignments, and no probability assignment whatsoever. They correspond in turn to each of the horns in the Four Horn theorem, which thus unifies them.

Consider again the 'no probability assignment whatsoever' trouble spots. I have argued that the probability of A given B can be defined without $P(A \cap B)$ or $P(B)$ being defined, so we should reject the ratio analysis. But you may wonder whether my argument proves too much. Can't we run a parallel argument against virtually every theorem of probability? For example, a rational agent can assign $P(A)$ a well-defined value, without assigning any values whatsoever to $P(A \cap B)$ and to $P(A \cap \neg B)$. Shouldn't we then reject the theorem of probability that P(A) = $P(A \cap B) + P(A \cap \neg B)$? After all, there are cases in which the left-hand side is defined, the right-hand side is undefined, and 'defined' does not equal 'undefined'![26]

This parallel argument does not prove too much; rather, it proves exactly enough: $P(A \cap B) + P(A \cap \neg B)$ is no *analysis* of $P(A)$. What this theorem, and all the other theorems of probability provide, are *constraints*: when all the terms that appear in the theorem are defined, they must conform to that theorem. If, on the other hand, some of the terms are undefined, then what we have is not a violation of the theorem, but rather a *non-instance* – something that simply does not fall under the theorem.

Perhaps the relevant lesson carries over to conditional probability: while the ratio does not provide an *analysis* of conditional probability, perhaps it does provide a *constraint* on rational opinion and on chance. It tells us is that whenever $P(A \cap B)$ and $P(B)$ are sharply defined, and $P(B)$ is non-zero, then the probability of A given B is constrained to be their ratio.

My attack on the ratio, seen as an analysis, exploited examples in which these conditions failed to be met. Nevertheless, the ratio might be thought of as a successful *partial* analysis, one that works for an important sub-class of conditional probabilities, in which the conditions *are* met. A sufficient condition for a conditional probability to equal a particular value is for the corresponding ratio to equal that value. However, it is not a necessary condition: a conditional probability can equal a particular value *without* the corresponding ratio equaling that value. At best, this leaves unfinished the project of giving a correct analysis of conditional probability (one that would seem to be particularly pressing in the cases of the Born rule, probabilistic causation, and so on).

But perhaps the very project of analyzing conditional probability was misguided from the start.

10. WHAT CONDITIONAL PROBABILITY *COULD* BE

A few years ago, a television commercial for Twix bars showed a schoolteacher in front of a bored class. The teacher asked the students: "Who can tell me the formula for conditional probability?" (Nobody answered – they ate Twix bars instead.) I have argued that the ratio formula does not give an adequate answer to this question. The bulk of this paper thus far has been negative; I want now to conclude more positively.

The various examples that I have considered show that often assignments of conditional probability are basic and immediate. They can, and in some cases must, stand without support from corresponding unconditional probabilities. We saw this in examples based on intuition, on scientific practice, and philosophical practice. Time and time again we found ourselves endorsing various intuitions, or scientific or philosophical theories, that committed us directly to specific values or constraints on conditional probabilities, apparently in the absence of the putative unconditional probabilities.

Moreover, the examples of vague and undefined probabilities suggest that the problem with the ratio analysis is not that it is a *ratio* analysis, as opposed to some other function of unconditional probabilities. The problem lies in the very attempt to analyze conditional probabilities in terms of unconditional probabilities at all. It seems that any other putative analysis that treated unconditional probability as more basic than conditional probability would meet a similar fate – as Kolmogorov's elaboration (RV) did.

On the other hand, given an unconditional probability, there is always a corresponding conditional probability lurking in the background. Your assignment of 1/2 to the coin landing heads superficially seems unconditional; but really it is conditional on tacit assumptions about the coin, the toss, the immediate environment, and so on. In fact, it is conditional on your total evidence. Furthermore, we can be sure that there can be no analogue of my argument, that conditional probabilities can be defined even when the corresponding unconditional probabilities are not, that runs the other way. For whenever an unconditional probability $P(X)$ is defined, it trivially equals the conditional probability $P(X$, given $\mathbf{T})$, where $\mathbf{T}$ is necessary. Unconditional probabilities are special cases of conditional probabilities.

Putting these facts together, we should regard conditional probability as conceptually prior to unconditional probability. So I suggest that we reverse the traditional direction of analysis: regard conditional probability to be the primitive notion, and unconditional probability as the derivative notion. I am effectively advocating a return to the views of some of the great figures in the philosophy of probability from yesteryear: among others, Johnson (1921), Keynes (1921), Carnap (1952), Jeffreys (1961), and de Finetti (1974, reprinted 1990). Their insights seem to have been overlooked or forgotten by many along the way, partly thanks to the hegemony of Kolmogorov's approach, I daresay. Indeed, de Finetti (1990, p. 134) forcefully contends not only that conditional probability is the primitive notion, but that in an important sense, *all* probability is conditional: "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".

What's more, thanks to authors such as Popper (1959b, 1992), Renyi (1970), van Fraassen (1995b), Spohn (1986) and Roeper and Leblanc (1999) we have axiomatizations of conditional probability ready to hand (although I do not want to insist that they are the last words on the subject either). They were, I think, motivated particularly by the problem of conditions with probability zero, although I suggest that the problems of conditions with vague or undefined probabilities provide at least as much motivation. There is now no difficulty, for example, in upholding the self-evident truth that the conditional probability of any (non-empty) proposition, given itself, is 1 – even for previously troublesome cases such as these. One can simply lay down an axiom to that effect.

Looking at Popper's system, the axiom of particular interest is the multiplication axiom, which in our notation is:

$$P(A \cap B,\ \text{given } C) = P(A,\ \text{given } B \cap C).P(B,\ \text{given } C).$$

(The other axiomatizations have analogous axioms.) Let C be necessary, and identify probability conditional on C with the corresponding unconditional probability. Then we have:

$$P(A \cap B) = P(A,\ \text{given } B).P(B).$$

That is, we recover the analogue of (PRODUCT) in Section 4.5. When $P(B) > 0$, of course, we can rearrange this to get the analogue of (RATIO). We recover what was right about the ratio analysis, while allowing us to go beyond it, giving the answers we want in the places where it breaks down.

So, returning to the question asked by the teacher in the Twix bar commercial, I want to conclude by rejecting the presupposition behind the question. There is no adequate formula for conditional probability, since it is the fundamental concept in probability theory.

ACKNOWLEDGEMENTS

I have many people to thank for many helpful discussions over many years: notable among them, Horacio Arló-Costa, Paul Benacerraf, Jeremy Butterfield, John Collins, Robin Collins, Fiona Cowie, James Hansen, Gil Harman, Allen Hazen, Christopher Hitchcock, Richard Jeffrey, Martin Jones, Martin Leckey, Renée Legare, Isaac Levi, Michaelis Michael, Ralph Miles, Daniel Nolan, Graham Priest, Mark Rigstad, Peter Vanderschraaf, and especially John Barker, Alex Byrne, Mark Colyvan, Adam Elga, Branden Fitelson, Ned Hall, Matthias Hild, Jim Joyce, Mark Kalderon, David Lewis, Ben Miller, Andrew Milne, Christian Piller, Teddy Seidenfeld, Brian Skyrms, Michael Thau, Bas van Fraassen, Peter Vranas, Jim Woodward (who also suggested the title), and Lyle Zynda, all of whom gave me incisive and in many cases extensive comments on various versions of this material. Another thank you to Michael Thau, who brilliantly fulfilled the functions of sounding-board, critic, and editor on several occasions. I also received helpful feedback from audiences at the Department of Computer Science of the Czech Academy of Sciences, the Australasian Artificial Intelligence Institute, the Department of Mathematics at USC, the Pittsburgh Center for Philosophy of Science conference in Bariloche, and the Departments of Philosophy at Auckland University, Victoria University, Canterbury University, Jagiellonian University, Johns Hopkins University, Cornell University, and MIT. I am grateful to the Whiting Foundation for financial support during the early stages of this project.

NOTES

1 Kolmogorov went on to develop a more sophisticated approach to conditional probability using the notion of conditional probability as a random variable. I will return to this approach in Section 4.5.

2 Modulo Kolmogorov's more sophisticated formulation of conditional probability, to be discussed later.

3 Analogies are easily multiplied. The material conditional $A \supset B$ is defined via the truth table for $\neg A \vee B$. But it would be specious to claim that the conditional 'if ... then' in English is *defined* via that truth table – whether or not '$\supset$' adequately analyzes 'if ... then' is a substantive question. Or simply consider *un*conditional probability. Kolmogorov, I said, gave us a certain axiomatization of the concept, but it would be specious to claim that his axioms are *definitive* of the concept. If they were, someone like de Finetti who disputes the axiomatization would simply appear foolish or ignorant, like someone disputing the definition of the material conditional. 'Conditional probability', like 'if ... then', like 'unconditional probability', and unlike 'material conditional' is a pre-theoretical concept of ours.

Despite the different subject matter, my comments echo ones made by Tarski (1956, 1983) in his seminal paper on logical consequence: "The concept of *logical consequence* is one of those whose introduction into the field of strict formal investigation was not a matter of arbitrary decision on the part of this or that investigator; ... efforts were made to adhere to the common usage of the language of everyday life" (1983, p. 409). I believe that the concept of conditional probability is at least as ensconced in "the language of everyday life" as is the concept of logical consequence. And I take my critique of the ratio analysis to be analogous in spirit to Etchemendy's (1990) critique of Tarski's analysis of logical consequence.

I should mention two allies in my campaign. Dorothy Edgington (1996, p. 621), writes in connection with the so-called "definition" of conditional probability: "Axiomatizations are not the life-blood of a theory. Their "definitions" have to fit pre-axiomatic working concepts". And Judea Pearl (1988, p. 70) contends that "Kolmogorov's axiomatization of probability is responsible for the unfortunate tradition of treating [(RATIO)] as a definition of conditional probability, rather than a theorem that follows from more primitive axioms about conditioning".

[4] For more on the need to distinguish subjective and objective probability, see for example Lewis (1980).

[5] This would certainly seem to be a live possibility on a Mill–Ramsey–Lewis style account of laws as regularities that appear as theorems in a 'best' theory of the universe, as long as the criteria for what makes one theory better than another are themselves vague. (In Lewis' 1973 theory, for instance, the vagueness may enter in the standards for balancing the theoretical virtues of 'simplicity' and 'strength'.) Then nature may not determine a single best theory, but rather a multiplicity of such theories. Suppose, for example, that these equal-best theories disagree on the chance that a radium atom decays in 1500 years: for each real number r in the interval [1/3, 2/3], there is such a theory that says that the chance is r. Then we might say that the chance of this event is vague over this interval.

There may be even more straightforward ways for the laws of nature to be vague. Perhaps some of the fundamental physical constants are not entirely precise – perhaps, for example, the gravitational constant is only fixed up to 100 decimal places. Or some of the fundamental physical properties might be vague. It would seem that the laws in which such constants or properties figure would then be rendered vague.

[6] A more complete introduction to non-standard analysis and infinitesimal probabilities can be found in Skyrms (1980) and (1995). My discussion here follows his (1995).

[7] De Finetti (1972) imagines regarding each of a denumerable collection of rival hypotheses as equally probable. He argues that each hypothesis should receive probability 0, yet the disjunction of all of them should receive probability 1. This is thus at once an argument against countable additivity, and against regularity; we will later see it providing an argument against the ratio analysis as well. Levi (1978) considers a continuum of hypotheses h_p, each asserting that the chance that a given coin lands heads is p, for each p in [0, 1]. Suppose an agent's state of opinion is represented by a continuous density function over the various possible values of p. Then the agent assigns probability 0 to each value of p, yet regards each value as a serious possibility. Here countable additivity can be obeyed, even though regularity is violated; and again, Levi turns it into an argument against what I am calling the ratio analysis.

[8] I am grateful to Matthias Hild, Christopher Hitchcock, Ben Miller and Peter Vranas, whose comments helped me streamline this section.

[9] See for instance de Finetti (1972) and Levi (1978). Return to their examples in footnote 7. In de Finetti's example, let E be a disjunction of some finite collection of the hypotheses, and let H be a member of this collection. $P(H$, given $E)$ should be defined, even though $P(E) = 0$. In Levi's example, P(coin lands heads, given h_p) should be p, even though $P(h_p) = 0$.

I hasten to add that the zero denominator problem is a problem for the ratio analysis of conditional probability, and not for Kolmogorov's more sophisticated analysis – see Section 4.5 – which tackles the problem head-on. It will face other problems.

[10] I thank Michael Thau for this point.

[11] This elaboration is not without its critics, though. It *requires* P to be countably, and not merely finitely, additive, so it is not welcomed by critics of countable additivity such as de Finetti. Moreover, there are infinitely many candidates for the name '$P(A \parallel \mathcal{F})$', for if the random variable f is one and $P(g = f) = 1$ (g and f agree 'almost everywhere'), then g is another. So the notation '$P(A \parallel \mathcal{F})$' merely denotes some one of the functions that satisfy (RV). Finally, getting the 'right' answers for the conditional probabilities requires one to choose the sub-σ-fields judiciously, and different choices can yield different answers to the same question.

[12] There are exceptions. A notable one is Seidenfeld (e.g., 2001), who has studied some of the ramifications of the analysis in detail. See also Seidenfeld et al. (2001).

[13] I am grateful to Adam Elga, Chris Hitchcock, Wilhelm Luxemburg, Vann McGee, Ben Miller, Brian Skyrms, and Michael Thau for very helpful discussion here.

[14] Modulo the caveat that one might not think that a function that violates countable additivity deserves the name 'probability function'.

[15] I have in mind here the infinitesimals of non-standard analysis, à la Robinson. I admit that this leaves open the possibility that some *other* construction of infinitesimals might solve this problem.

[16] Paralleling my caveat regarding infinitesimals in the previous footnote, I admit that my argument leaves open whether *other* ways of modeling vague probability are susceptible to the same objections.

[17] I thank Michael Thau for pointing this out to me.

[18] Price (1986), Mellor (1993), and Edgington (1995) all make versions of the point that there are cases in which conditional probabilities are defined while the corresponding unconditional probabilities happen not to be defined. But they make this point only for *subjective probability*. And I will make the stronger point: there are such cases in which the corresponding unconditional probabilities *cannot* be defined. I will discuss their arguments further in Sections 8.3 and 8.4.

[19] I take it that it is just such a position that Price (1986) considers (and rejects, though for different reasons to mine), namely: "...thinking of a conditional credence as a full belief about the ratio ... On this view a person might hold the conditional credence – the belief about the ratio – without the associated absolute credences. It would simply be like believing that one thing was twice as long as another without knowing the length of each" (20).

[20] I am grateful to Adam Elga for this point.

[21] Exact formulations vary, but not in ways that count against my claim here. Some authors restrict the condition M to the choice of measurement, and index the probability function according to the state of the system. For example, Jarrett (1989, p. 66) considers conditional probabilities in a Bell experiment of the form $P_A^\lambda(x \mid i, j, y)$, where λ is the state of the two particle system, x and y are the outcomes, and i and j are the switch settings at the ends of the experiment (of which A is one). Here i and j correspond to choices of measurements of certain spin observables, so they play the same role as M does in my formulation. For what I take to be similar allegiance to the ratio analysis, see for example Fine (1989) and Howard (1989).

Also, it is standard practice to express quantum mechanics' *transition probabilities* in terms of the ratio analysis (see e.g. van Fraassen 1991, p. 117). I think that this practice is equally questionable.

[22] When I speak of "quantum mechanics itself", I mean the uninterpreted theory – that which all interpretations of quantum mechanics agree on.

[23] A measurement is typically performed by some experimenter manipulating some piece of apparatus – although my arguments will assume only that *some* measurements are like this.

[24] Here I am grateful to Ned Hall for helpful discussion.

[25] There are subtleties that I cannot go into here, including the notion of admissibility, the relativity of chances to times, and Lewis' (1994b) revised version of the Principle. These subtleties will not affect my main point.

[26] I thank John Barker and (independently) Christian Piller for putting this objection to me.

REFERENCES

Appiah, Anthony: 1985, *Assertion and Conditionals*, Cambridge University Press, Cambridge.

Carnap, Rudolf: 1952, *The Continuum of Inductive Methods*, University of Chicago Press.

Carnap, Rudolf: 1963, 'Carnap's Intellectual Autobiography', in Paul Arthur Schilpp (ed.), *The Philosophy of Rudolf Carnap*, The Library of Living Philosophers, Vol. XI.

Cartwright, Nancy: 1979, 'Causal Laws and Effective Strategies', *Noûs* **13**, 419–437.

Cartwright, Nancy: 1999, *The Dappled World*, Cambridge University Press, Cambridge.

Cushing, James T. and Ernan McMullin (eds.): 1989, *Philosophical Consequences of Quantum Theory – Reflections on Bell's Theorem*, University of Notre Dame Press, Indiana, pp. 224–253.

de Finetti, Bruno: 1972, *Probability*, *Induction and Statistics*, Wiley, New York.

de Finetti, Bruno: 1980, 'Foresight: Its Logical Laws, Its Subjective Sources', in Henry E. Kyburg, Jr. and Howard E. Smokler (eds.), *Studies in Subjective Probability*, Robert E. Krieger Publishing Company, pp. 53–118.

de Finetti, Bruno: 1990 (originally published 1974), *Theory of Probability*, Vol. 1, Wiley Classics Library, John Wiley & Sons.

Earman, John: 1992, *Bayes or Bust*?, MIT Press.

Edgington, Dorothy: 1995, 'On Conditionals', *Mind* **104**, 235–329.

Edgington, Dorothy: 1996, 'Lowe on Conditional Probability', *Mind* **105**, 617–630.

Edwards, W., Lindman, H., and Savage, L. J.: 1963, 'Bayesian Statistical Inference for Psychological Research', *Psychological Review* **LXX**, 193–242.

Eells, Ellery: 1991, *Probabilistic Causality*, Cambridge University Press, Cambridge.

Etchemendy, John: 1990, *The Concept of Logical Consequence*, Harvard University Press.

Fetzer, James: 1983, 'Probability and Objectivity in Deterministic and Indeterministic Situations', *Synthese* **57**, 367–386.

Fine, Arthur: 1989, 'Do Correlations Need to be Explained?', in Cushing and McMullin, pp. 175–194.

Fine, Terrence: 1973, *Theories of Probability*, Academic Press.

Gaifman, Haim:1988, 'A Theory of Higher Order Probabilities', in Brian Skyrms and William L. Harper (eds.), *Causation*, *Chance*, *and Credence*, Kluwer Academic Publishers, pp. 191–219.

Gillies, Donald: 2000, 'Varieties of Propensity', *British Journal for the Philosophy of Science* **51**, 807-835.

Gilboa, Itzhak: 1994, 'Can Free Choice Be Known?', in C. Bicchieri, R. Jeffrey and B. Skyrms (eds.), *The Logic of Strategy*, Oxford University Press.

Goldstein, M.: 1983, 'The Prevision of a Prevision', *Journal of the American Statistical Association* **78**, 817–819.

Hacking, Ian: 1965, *The Logic of Statistical Inference*, Cambridge University Press, Cambridge.

Hájek, Alan: 1997, ' "*Mises Redux*" – *Redux*: Fifteen Arguments Against Finite Frequentism', *Erkenntnis* **45**, 209–227.

Howard, Don: 1989, 'Holism, Separability and the Metaphysical Implications of the Bell Experiments', in Cushing and McMullin.

Howson, Colin and Peter Urbach: 1993, *Scientific Reasoning: The Bayesian Approach*, Open Court, 2nd edition.

Jackson, Frank: 1987, *Conditionals*, Blackwell.

Jarrett, Jon P.: 1989, 'Bell's Theorem: A Guide to the Implications', in Cushing and McMullin, pp. 60–79.

Jeffrey, Richard: 1983a, 'Bayesianism with a Human Face', in John Earman (ed.), *Testing Scientific Theories*, Minnesota Studies in the Philosophy of Science, vol. X, pp. 133–156.

Jeffrey, Richard: 1983b, *The Logic of Decision*, 2nd ed., University of Chicago Press.

Jeffreys, Harold: 1961, *Theory of Probability*, 3rd ed., Oxford, Clarendon Press.

Johnson, W. E.: 1921, *Logic*, Cambridge University Press.

Keynes, John Maynard: 1921, *A Treatise on Probability*, Macmillan and co.

Kemeny, J.: 1955, 'Fair Bets and Inductive Probabilities', *Journal of Symbolic Logic* **20**, 263–273.

Kolmogorov, A. N.: 1933, *Grundbegriffe der Wahrscheinlichkeitrechnung*, Ergebnisse Der Mathematik; translated as *Foundations of Probability*, Chelsea Publishing Company, 1950.

Kyburg, Henry: 1988, 'Powers', in William L. Harper and Brian Skyrms (eds.), *Causation in Decision*, *Belief Change*, *and Statistics*, Kluwer Academic Publishers, pp. 71–82.

Levi, Isaac: 1974, 'On Indeterminate Probabilities', *The Journal of Philosophy* **71**, 391–418.

Levi, Isaac: 1978, 'Coherence, Regularity and Conditional Probability', *Theory and Decision* **9**, 1–15.

Levi, Isaac: 1980, *The Enterprise of Knowledge*, The MIT Press.

Levi, Isaac: 1997, *The Covenant of Reason: Rationality and the Commitments of Thought*, Cambridge University Press.

Lewis, David: 1973, *Counterfactuals*, Harvard University Press, Cambridge, Massachusetts.

Lewis, David: 1980, 'A Subjectivist's Guide to Objective Chance', in *Studies in Inductive Logic and Probability*, Vol. II., University of California Press, pp. 263–293; reprinted in Lewis 1986b.

Lewis, David: 1986a, 'Probabilities of Conditionals and Conditional Probabilities II', *Philosophical Review* **95**, 581–589.

Lewis, David: 1986b, *Philosophical Papers Volume II*, Oxford University Press.

Lewis, David: 1986c, *On the Plurality of Worlds*, Blackwell.

Lewis, David: 1994a, 'Reduction of Mind', in Samuel Guttenplan (ed.), *A Companion to the Philosophy of Mind*, Blackwell, pp. 412–431.

Lewis, David: 1994b, 'Humean Supervenience Debugged', *Mind* **103**, 473–490.

Lindstrøm, Tom: 1988, 'An Invitation to Nonstandard Analysis', in Nigel Cutland (ed.), *Nonstandard Analysis and its Applications*, pp. 1–105.

Mellor, D. H.: 1971, *The Matter of Chance*, Cambridge University Press, Cambridge.

Mellor, D. H.: 1993, 'How to Believe a Conditional', *Journal of Philosophy* **90**, 233–248.

Pearl, Judea: 1988, *Probabilistic Reasoning in Intelligent Systems*, Morgan-Kaufman.

Pitowski, Itamar: 1989, *Quantum Probability – Quantum Logic*, Springer-Verlag.

Popper, Karl: 1959a, 'The Propensity Interpretation of Probability', *British Journal of the Philosophy of Science* **10**, 25–42.

Popper, Karl: 1959b, *The Logic of Scientific Discovery*, Basic Books; reprint edition 1992, Routledge.

Price, Huw: 1986, 'Conditional Credence', *Mind* **95**, 18–36.

Ramsey, F. P.: 1926, 'Truth and Probability', in R. B. Braithwaite (ed.), *Foundations of Mathematics and other Essays*, Routledge & Kegan Paul, 1931, pp. 156–198; reprinted in Henry E. Kyburg, Jr. and Howard E. Smokler (eds.), *Studies in Subjective Probability*, 2nd ed., Robert E. Krieger Publishing Company, 1980, pp. 23–52; reprinted in D. H. Mellor (ed.), *Philosophical Papers*, Cambridge University Press, Cambridge, 1990, pp. 52–96.

Reichenbach, Hans: 1949, *The Theory of Probability*, University of California Press, Berkeley.

Renyi, Alfred: 1970, *Foundations of Probability*, Holden-Day, Inc.

Robinson, Abraham: 1966, *Non-Standard Analysis*, North Holland Publishing Company, Amsterdam.

Roeper, P. and H. Leblanc: 1999, *Probability Theory and Probability Logic*, University of Toronto Press.

Rosenkrantz, R. D.: 1981, *Foundations and Applications of Inductive Probability*, Ridgeview Publishing Company.

Salmon, Wesley: 1980, 'Probabilistic Causality', *Pacific Philosophical Quarterly* **61**, 50–74.

Savage, L. J.: 1954, *The Foundations of Statistics*, John Wiley.

Schervish, M., T. Seidenfeld, and J. Kadane: 1984, 'The Extent of Non-Conglomerability of Finitely Additive Probability', *Z. War.* **66**, 205–226.

Seidenfeld, T.: 2001, 'Remarks on the Theory of Conditional Probability: Some Issues of Finite Versus Countable Additivity', in V. F. Hendricks et al. (eds.), *Probability Theory*, Kluwer Academic Publishers, pp. 167–178.

Seidenfeld, Teddy, Mark J. Schervish, and Joseph B. Kadane: 2001, 'Improper Regular Conditional Distributions', *The Annals of Probability* **29**(4), 1612–1624.

Shimony, Abner: 1970, 'Scientific Inference', in Robert Colodny (ed.), *The Nature and Function of Scientific Theories*, University of Pittsburgh Press.

Skyrms, Brian: 1980, *Causal Necessity*, Yale University Press.

Skyrms, Brian: 1995, 'Strict Coherence, Sigma Coherence and the Metaphysics of Quantity', *Philosophical Studies* **77**, 39–56.

Skyrms, Brian: 2000, *Choice and Chance*, Wadsworth, Inc, 4th ed.

Spohn, Wolfgang: 1977, 'Where Luce and Krantz Do Really Generalize Savage's Decision Model', *Erkenntnis* **11**, 113–134.

Spohn, Wolfgang: 1986, 'The Representation of Popper Measures', *Topoi* **5**, 69–74.

Stalnaker, Robert: 1970, 'Probabilities and Conditionals', *Philosophy of Science* **37**, 64–80.

Suppes, Patrick: 1970, *A Probabilistic Theory of Causality*, North Holland Publishing Company, Amsterdam.

Tarski, Alfred: 1956, 'On the Concept of Logical Consequence', in *Logic*, *Semantics*, *Metamathematics*, Oxford: Clarendon Press; Hackett Publishing Company, Indianapolis, Indiana, 1983. (Trans. by J. H. Woodger, 2nd. ed. edited and introduced by John Corcoran.)

van Fraassen, Bas: 1984, 'Belief and the Will', *Journal of Philosophy* **81**, 235–256.

van Fraassen, Bas: 1989, *Laws and Symmetry*, Clarendon Press, Oxford.

van Fraassen, Bas: 1990, 'Figures in a Probability Landscape', in J. M. Dunn and A. Gupta (eds.), *Truth or Consequences*, Kluwer, pp. 345–356.

van Fraassen, Bas: 1991, *Quantum Mechanics – An Empiricist View*, Clarendon Press, Oxford.

van Fraassen, Bas: 1995a, 'Belief and the Problem of Ulysses and the Sirens', *Philosophical Studies* **77**, 7–37.

van Fraassen, Bas: 1995b, 'Fine-Grained Opinion, Conditional Probability, and the Logic of Belief', *Journal of Philosophical Logic* **24**, 349–377.

Humanities and Social Sciences 101-40
Caltech, Pasadena, CA 91125, USA
E-mail: ahajek@hss.caltech.edu