CONDITIONAL PROBABILITY

Alan Hájek

1 INTRODUCTION

A fair die is about to be tossed. The probability that it lands with '5' showing up is 1/6; this is an *unconditional probability.* But the probability that it lands with '5' showing up, *given* that it lands with an odd number showing up, is 1/3; this is a *conditional probability.* In general, conditional probability is probability *given* some body of evidence or information, probability *relativised to* a specified set of outcomes, where typically this set does not exhaust all possible outcomes. Yet understood that way, it might seem that *all* probability is conditional probability — after all, whenever we model a situation probabilistically, we must initially delimit the set of outcomes that we are prepared to countenance. When our model says that the die may land with an outcome from the set $\{1, 2, 3, 4, 5, 6\}$, it has already ruled out its landing on an edge, or on a corner, or flying away, or disintegrating, or ..., so there is a good sense in which it is taking the non-occurrence of such anomalous outcomes as "given". Conditional probabilities, then, are supposed to earn their keep when the evidence or information that is "given" is more specific than what is captured by our initial set of outcomes. In this article we will explore various approaches to conditional probability, canvassing their associated mathematical and philosophical problems and numerous applications. Having done so, we will be in a better position to assess whether conditional probability can rightfully be regarded as the fundamental notion in probability theory after all.

Historically, a number of writers in the pantheon of probability took it to be so. Johnson [1921], Keynes [1921], Carnap [1952], Popper [1959b], Jeffreys [1961], Renyi [1970], and de Finetti [1974/1990] all regarded conditional probabilities as primitive. Indeed, de Finetti [1990, 134] went so far as to say that "every prevision, and, in particular, every evaluation of probability, is conditional; not only on the mentality or psychology of the individual involved, at the time in question, but also, and especially, on the state of information in which he finds himself at that moment". On the other hand, orthodox probability theory, as axiomatized by Kolmogorov [1933], takes unconditional probabilities as primitive and later analyses conditional probabilities in terms of them. Whatever we make of the primacy, or otherwise, of conditional probability, there is no denying its importance, both in probability theory and in the myriad applications thereof — so much so that the author of an article such as this faces hard choices of prioritisation. My choices are targeted more towards a philosophical audience, although I hope that they will be of wider interest as well.

Handbook of the Philosophy of Science. Volume 7: Philosophy of Statistics.
Volume editors: Prasanta S. Bandyopadhyay and Malcolm R. Forster. General Editors: Dov M. Gabbay, Paul Thagard and John Woods.

2 MATHEMATICAL THEORY

2.1 Kolmogorov's axiomatization, and the ratio formula

We begin by reviewing Kolmogorov's approach. Let Ω be a non-empty set. A *field* (*algebra*) on Ω is a set $\mathcal{F}$ of subsets of Ω that has Ω as a member, and that is closed under complementation (with respect to Ω) and union. Assume for now that $\mathcal{F}$ is finite. Let P be a function from $\mathcal{F}$ to the real numbers obeying:

1. $P(A) \geq 0$ for all $A \in \mathcal{F}$. (Non-negativity)
2. $P(\Omega) = 1$. (Normalization)
3. $P(A \cup B) = P(A) + P(B)$ for all $A, B \in \mathcal{F}$ such that $A \cap B = \emptyset$ (Finite additivity)

Call P a *probability function*, and $(\Omega, \mathcal{F}, P)$ a *probability space*.

One could instead attach probabilities to members of a collection of *sentences* of a formal language, closed under truth-functional combinations.

Kolmogorov extends his axiomatization to cover infinite probability spaces. Probabilities are now defined on a *σ-field* (*σ-algebra*) — a field that is further closed under *countable* unions — and the third axiom is correspondingly strengthened:

3′. If $A_1, A_2, \ldots$ is a countable sequence of (pairwise) disjoint sets, each belonging to $\mathcal{F}$, then
$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$ (Countable additivity)

So far, all probabilities have been unconditional. Kolmogorov then introduces the *conditional probability of A given B* as the ratio of unconditional probabilities:

(RATIO) $$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ provided } P(B) > 0$$

(On the sentential formulation this becomes:

$$P(A|B) = \frac{P(A\&B)}{P(B)}, \text{ provided } P(B) > 0.)$$

This is often called the "definition" of conditional probability, although I suggest that instead we call it a *conceptual analysis*[1] of conditional probability. For 'conditional probability' is not simply a technical term that one is free to introduce however one likes. Rather, it begins as a pre-theoretical notion for which we have

[1]Or (prompted by Carnap [1950] and Maher [2007]), perhaps it is an *explication*. I don't want to fuss over the viability of the analytic/synthetic distinction, and the extent to which we should be refining a folk-concept that may not even be entirely coherent. Either way, my point stands that Kolmogorov's formula is not merely definitional.

associated intuitions, and Kolmogorov's ratio formula is answerable to those. So while we are free to stipulate that '$P(A|B)$' is merely shorthand for this ratio, we are *not* free to stipulate that 'the conditional probability of A, given B' should be identified with this ratio. Compare: while we are free to stipulate that '$A \supset B$' is merely shorthand for a connective with a particular truth table, we are *not* free to stipulate that 'if A, then B' in English should be identified with this connective.

And Kolmogorov's ratio formula apparently answers to most of our intuitions wonderfully well.

2.2 *Support for the ratio formula*

Firstly, it is apparently supported on a case-by-case basis. Consider the fair die example. Intuitively, the probability of '5', given 'odd', is 1/3 because we imagine narrowing down the possible outcomes to the three odd ones, observing that '5' is one of them and that probability is shared equally among them. And the ratio formula delivers this verdict:

$$P(5|odd) = \frac{P(5 \cap \text{odd})}{P(\text{odd})} = \frac{1/6}{1/2} = 1/3.$$

And so it goes with countless other examples.

Secondly, a nice heuristic for Kolmogorov's axiomatization is given by van Fraassen's [1989] "muddy Venn diagram" approach, which suggests an informal argument in favour of the ratio formula. Think of the usual Venn-style representation of sets as regions inside a box (depicting Ω). Think of probability as mud spread over the diagram, so that the amount of mud sitting above a given region corresponds to its probability, with a total amount of 1 unit of mud. When we consider the conditional probability of A, given B, we restrict our attention to the mud that sits above the region representing B, and then ask what proportion of *that* mud sits above A. But that is simply the amount of mud sitting above $A \cap B$, divided by the amount of mud sitting above B.

Thirdly, the ratio formula can be given a frequentist justification. Suppose that we run a long sequence of n trials, on each of which B might occur or not. It is natural to identify the probability of B with the relative frequency of trials on which it occurs:

$$P(B) = \frac{\#(B)}{n}$$

Now consider among *those* trials the proportion of those on which A also occurs:

$$P(A|B) = \frac{\#(A \cap B)}{\#(B)}$$

But this is the same as

$$\frac{\#(A \cap B)/n}{\#(B)/n}$$

which on the frequentist interpretation is identified with

$$\frac{P(A \cap B)}{P(B)}.$$

Fourthly, the ratio formula for subjective conditional probability is supported by an elegant *Dutch Book* argument originally due to de Finetti [1937] (here simplified). Begin by identifying your subjective probabilities, or *credences*, with your corresponding betting prices. You assign probability p to X if and only if you regard pS as the value of a bet that pays S if X, and nothing otherwise. Symbolize this bet as:

S	if X
0	otherwise

For example, my credence that this coin toss results in heads is $1/2$, corresponding to my valuing the following bet at 50 cents:

\$ 1	if heads
0	otherwise

A *Dutch Book* is a set of bets bought or sold at such prices as to guarantee a net loss. An agent is susceptible to a Dutch Book if there exists such a set of bets, bought or sold at prices that she deems acceptable. Now introduce the notion of a *conditional bet* on A, given B, which

- pays \$1 if $A \cap B$
- pays 0 if $A^c \cap B$
- is called off if B^c (that is, the price you pay for the bet is refunded if B does not occur).

Identify your $P(A|B)$ with the value you attach to this conditional bet — that is, to:

\$1	if $A \cap B$
0	if $A^c \cap B$
$P(A\|B)$	if B^c

Now we can show that if your credences violate (RATIO), you are susceptible to a Dutch Book. For the conditional bet can be regarded as equivalent (giving the same pay-offs in every possible outcome) to the following pair of bets:

\$ 1	if $A \cap B$
0	if $A^c \cap B$
0	if B^c

0	if $A \cap B$
0	if $A^c \cap B$
$P(A\|B)$	if B^c

which you value at $P(A \cap B)$ and $P(A|B)P(B^c)$ respectively. So to avoid being Dutch Booked, you must value the conditional bet at $P(A \cap B) + P(A|B)P(B^c)$. That is:

$$P(A|B) = P(A \cap B) + P(A|B)P(B^c),$$

from which the ratio formula follows (since $P(B^c) = 1 - P(B)$).

2.3 Some basic theorems involving conditional probability

With Kolmogorov's axioms and ratio formula in place, we may go on to prove a number of theorems involving conditional probability. Especially important is the *law of total probability*, the simplest form of which is:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

This follows immediately from the additivity formula

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

by two uses of (RATIO). The law generalizes to the case in which we have a countable partition $B_1, B_2, \ldots$:

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots$$

This tells us that the unconditional probability $P(A)$ can be identified with a weighted average, or expectation, of probabilities conditional on each cell of a partition, the weights being the unconditional probabilities of the cells. We will see how the theorem underpins Kolmogorov's more sophisticated formulation of conditional probability (§5), and a rule for updating probabilities (Jeffrey conditionalization, §7.2). In the meantime, notice how it yields versions of the equally celebrated *Bayes' theorem:*

$$\begin{aligned} P(A|B) &= \frac{P(B|A)P(A)}{P(B)} && \text{(by two uses of (RATIO))} \\ &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} && \text{(by the law of total probability)} \end{aligned}$$

More generally, suppose there is a partition of hypotheses $\{H_1, H_2, ...\}$, and evidence E. Then for each i,

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_j P(E|H_j)P(H_j)}$$

The $P(E|H_i)$ terms are called *likelihoods.*

Bayes' theorem has achieved such a mythic status that an entire philosophical and statistical movement — "Bayesianism" — is named after it. This is meant to honour the important role played by Bayes' theorem in calculating terms of the form '$P(H_i|E)$', and (at least among philosophers) Bayesianism is particularly associated with a subjectivist interpretation of 'P', and correspondingly with the thesis that rational degrees of belief are probabilities. This may seem somewhat curious, as Bayes' theorem is neutral vis-à-vis the interpretation of probability, being purely a theorem of the formal calculus, and just one of many theorems at that. In particular, it provides just one way to calculate a conditional probability, when various others are available, all ultimately deriving from (RATIO); and as we will see, often conditional probabilities can be ascertained directly, without any calculation at all. Moreover, a diachronic prescription for *revising* or *updating* probabilities, which we will later call 'conditionalization', is sometimes wrongly called 'updating by Bayes' theorem'. In fact the theorem is a 'static' rule relating probabilities synchronically, and being purely a piece of mathematics it cannot by itself have any normative force.

2.4 Independence

Kolmogorov's axioms assimilate probability theory to measure theory, the general theory of length, area, and volume. (Think of how these quantities are non-negative, additive, and can often be normalized.) Conditional probability is a further, distinctively probabilistic notion without any obvious counterpart in measure theory. Similarly, *independence* is distinctively probabilistic, and ultimately parasitic on conditional probability. Let $P(A)$ and $P(B)$ both be positive. According to Kolmogorov's theory, A and B are *independent*

$$\begin{aligned} &\text{iff} \quad P(A|B) = P(A); \text{equivalently,} \\ &\text{iff} \quad P(B|A) = P(B). \end{aligned}$$

These equations are supposed to capture the idea of A and B being uninformative regarding each other: that one event occurs in no way affects the probability that the other does. To be sure, there is a further equivalent characterization of independence that is free of conditional probability:

$$P(A \cap B) = P(A)P(B).$$

But its rationale comes from the equalities of the corresponding conditional and unconditional probabilities. It has the putative advantage of applying even when A and/or B have probability 0; although it is questionable whether probability 0 events should automatically be independent of everything (including themselves, and their complements!).

When we say that 'A is independent of B', we suppress the fact that such independence is really a *three*-place relation between an event, an event, and *a probability function.* This distinguishes probabilistic independence from such two-place relations as logical and counterfactual independence. Probabilistic independence is assumed in many of probability theory's classic limit theorems.

We may go on to give a Kolmogorovian analysis of *conditional independence.* We say that A is independent of B, given C, if

$$P(A \cap B|C) = P(A|C)P(B|C), \text{ provided } P(C) > 0.$$

2.5 Conditional expectation

Conditional probability underlies the concept of *conditional expectation,* also important in probability theory. A *random variable* X (on Ω) is a function from Ω to the set of real numbers, which takes the value $X(\omega)$ at each point $\omega \in \Omega$. If X is a random variable that takes the values $x_1, x_2, \ldots$ with probabilities $p(x_1), p(x_2)$, $\ldots$, then the *expected value* of X is defined as

$$E(X) = \sum_i x_i p(x_i)$$

provided that the series converges absolutely. (For continuous random variables, we replace the $p(x_i)$ by values given by a density function, and replace the sum by an integral.) A *conditional expectation* is an expectation of a random variable with respect to a conditional probability distribution. Let X and Y be two random variables with joint distribution

$$P(X = x_j \cap Y = y_k) = p(x_j, y_k)(j, k = 1, 2, \ldots)$$

The conditional expectation $E(Y|X = x_j)$ of Y given $X = x_j$ is given by:

$$\sum_k y_k P(Y = y_k|X = x_j) = \frac{\sum_k y_k p(x_j, y_k)}{p(x_j)}$$

(Again, this has a continuous version.) We may generalize to conditional expectations involving more random variables.

So the importance of conditional probability in probability theory is beyond dispute. The same can be said of its role in many philosophical applications.

3 PHILOSOPHICAL APPLICATIONS

Conditional probability is near-ubiquitous in both the methodology — in particular, the use of statistics and game theory — of the sciences and social sciences, and in their specific theories. Various central concepts in statistics are defined in terms of conditional probabilities: significance level, power, sufficient statistics, ancillarity, maximum likelihood estimation, Fisher information, and so on. Game theorists appeal to conditional probabilities for calculating the expected payoffs in correlated equilibrium; computing the Bayesian equilibrium in games of incomplete information; in certain Bayesian dynamic updating models of equilibrium selection in repeated games; and so on. Born's rule in quantum mechanics is often understood as a method of calculating the conditional probability of a particular measurement outcome, given that a measurement of a certain kind is performed on a system in a certain state. In medical and clinical psychological testing, conditional probabilities of the form P(disorder | positive test) ("diagnosticity") and P(positive test | disorder) ("sensitivity") take centre stage. Mendelian genetics allows us to compute probabilities for an organism having various traits, given information about the traits of its parents; and population genetics allows us to compute the chance of a trait going to fixation, given information about population size, initial allele frequencies, and the fitness gradient. The regression equations of economics, and many of the results in time series analysis, are claims about conditional probabilities. And so it goes — this little sampler could be extended almost indefinitely.

Moreover, conditional probability is a staple of philosophy. The next section surveys a few of its philosophical applications.

3.1 Conditional probability in the philosophy of probability

A central issue in the philosophical foundations of probability is that of *interpreting* probability — that is, of analysing or explicating the 'P' that appears in its formal theory. Conditional probability finds an important place in all of the leading interpretations:

Frequentism: Probability is understood as relative frequency (perhaps in an infinite sequence of hypothetical trials) — e.g. the probability of heads for a coin is identified with the number of heads outcomes divided by the total number of trials in some suitable sequence of trials. Recalling our third justification for the ratio formula in §2.2, this seems to be naturally understood as a conditional probability, the condition being whatever determines the suitability of that sequence.

Propensity: Probability is a measure of the tendency for a certain kind of experimental set-up to produce a particular outcome, either in the single case [Giere, 1973], or in the long run [Popper, 1959a]. Either way, it is a conditional probability, the condition being a specification of the experimental set up.

Classical: Probability is assigned by one in an epistemically neutral position with respect to a set of "equally possible" cases — outcomes on which one's evidence bears equally. Such an assignment must thus be relativised to such evidence.

Logical: Probability is a measure of inductive support or partial entailment, generalizing both deductive logic's notion of entailment and the classical interpretation's assignments to "equally possible" cases. In Carnap's notation, $c(h, e)$ is a measure of the degree of support that evidence e confers on h. This is explicitly a conditional probability.

Subjective: Probability is understood as the degree of belief of some agent (typically assumed to be ideally rational). As we have seen, some subjectivists (e.g. Jeffreys, de Finetti) explicitly regarded subjective *conditional* probability to be basic. But even subjectivists who regard unconditional probability as basic find an important place for conditional probability. Subjectivists are unified in regarding conformity to the probability calculus as a rational requirement on credences. They often add further constraints, couched in terms of conditional probabilities; a number of examples follow.

Gaifman [1988] coins the term "expert probability" for a probability assignment that a given agent with subjective probability function P strives to track. We may codify this idea as follows (simplifying his characterization at the expense of some generality):

$$\text{(Expert) } P(A|pr(A) = x) = x, \text{ for all } x \text{ such that } P(pr(A) = x) > 0.$$

Here $pr(A)$ is the assignment that the agent regards as expert. For example, if you regard the local weather forecaster as an expert, and she assigns probability 0.1 to it raining tomorrow, then you may well follow suit:

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

More generally, we might speak of an entire probability function as being such a guide for an agent, over a specified set of propositions — so that (Expert) holds for any choice of A from that set. A *universal expert function* would guide *all* of the agent's probability assignments in this way. van Fraassen [1984; 1995], following Goldstein [1983], argues that an agent's *future* probability functions are universal expert functions for that agent — his *Reflection Principle*:

$$P_t(A|P_{t'}(A) = x) = x, \text{ for all } A \text{ and for all } x \text{ such that } P_t(P_{t'}(A) = x) > 0,$$

where P_t is the agent's probability function at time t, and $P_{t'}$ her function at later time t'. The principle encapsulates a certain demand for 'diachronic coherence' imposed by rationality. van Fraassen defends it with a 'diachronic' Dutch Book argument (one that considers bets placed at different times), 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. The *Principle of Direct Probability* regards the relative frequency function as a universal expert function (cf. [Hacking, 1965]). Let A be an event-type, and let $relfreq(A)$ be the relative frequency of A (in some suitable reference class). Then for any rational agent, we have

$$P(A|relfreq(A) = x) = x, \text{ for all } A \text{ and for all } x \text{ such that } P(relfreq(A) = x) > 0.$$

Related, but distinct according to those who do not identify objective chances with relative frequencies, is Lewis's [1980]:

$$\text{(Principal Principle) } P(A|ch_t(A) = x \,\&\, E) = x$$

Here 'P' is an 'initial' rational credence function (the prior probability function of a rational agent who has acquired no information), A is a proposition, $ch_t(A)$ is the chance of A at time t and E is further evidence that may be acquired. In order for the Principal Principle to be applicable, E cannot be relevant to whether A is true or false, other than by bearing on the chance of A at t; E is then said to be *admissible* (strictly speaking: with respect to P, A, t, and x). The literature ritually misstates the Principal Principle, regarding 'P' as the credence function of a rational agent quite generally, rather than an 'initial' credence function as Lewis explicitly formulated it. Misstated this way, it is open to easy counterexamples in which the agent has information bearing on A that has been incorporated into 'P', although not explicitly written into the condition (in the slot that 'E' occupies). Interestingly, admissibility is surely just as much an issue for the other expert principles, yet for some reason hardly discussed outside the Principal Principle literature, where it is all the rage.

Finally, some authors impose the requirement of *strict coherence* on rational agents: such an agent assigns $P(H|E) = 1$ only if E entails H. See Shimony [1955].

3.2 Some uses of conditional probability in other parts of philosophy

The use of conditional probability in updating rules for credences, and in the semantics of conditionals, has been so important and fertile that I will devote entire sections to them later on (§7 and §9). In the meantime, here are just a few of the myriad applications of conditional probability in various other areas of philosophy.

Probabilistic causation

A major recent industry in philosophy has been that of providing analyses of causation compatible with indeterminism. At a first pass, we might analyze 'causation' as 'correlation' — that is, analyze 'A causes B' as

$$P(B|A) > P(B|A^c).$$

This analysis cannot be right. It wrongly classifies spurious correlations and effects of common causes as instances of causation; moreover, it fails to capture the asymmetry of the causal relation. So a number of authors refine the analysis along the following lines (e.g. [Suppes, 1970; Cartwright, 1979; Salmon, 1980; Eells, 1991]):

> A causes B iff $P(B|A \cap X) > P(B|A^c \cap X)$ for every member X of some 'suitable' partition.

The exact details vary from author to author; what they share is the fundamental appeal to inequalities among conditional probabilities.

Reichenbach's [1956] famous *common cause principle* is again couched in terms of inequalities among conditional probabilities. The principle asserts that if A and B are simultaneous events that are correlated, then there exists an earlier common cause C of A and B, such that for every member X of some 'suitable' partition,

$$\begin{aligned}&P(A|C) > P(A|C^c),\\&P(B|C) > P(B|C^c),\\&P(A \cap X|C) = P(A|C)P(B|C), \text{ and}\\&P(A \cap X|C^c) = P(A|C^c)P(B|C^c).\end{aligned}$$

That is, C is correlated with A and with B, and C screens off A from B (they are independent conditional on C).

Bayesian networks

We may model a causal network as a directed acyclic graph with nodes corresponding to variables. If one variable directly causes another, we join the corresponding nodes with a directed edge, its arrow pointing towards the 'effect' variable. We may naturally employ a genealogical nomenclature. We call the cause the 'parent' variable, the effect a 'child' variable, and call iterations of these relationships 'ancestors' and 'descendants' in the obvious way.

In a Bayesian network, a probability distribution is assigned across the nodes. The *Causal Markov condition* is a commonly held assumption about conditional independence relationships. Roughly, it states that any node in a given network is conditionally independent of its non-descendents, given its parents. More formally (with obvious notation): "Let G be a causal graph with vertex set V and P be a probability distribution over the vertices in V generated by the causal structure represented by G. G and P satisfy the Causal Markov Condition if and only if for every W in V, W is independent of $V\backslash(\text{Descendants}(W) \cup \text{Parents}(W))$ given Parents(W)" [Spirtes *et al.*, 2000, 29]. ("\" denotes set subtraction.) *Faithfulness* is the converse condition that the set of independence relations derived from the Causal Markov Condition is exactly the set of independence relations that hold for the network. (See [Spirtes *et al.*, 2000; Hausman and Woodward, 1999].)

Inductive-statistical explanation:

Hempel [1965] regards scientific explanation as a matter of subsuming an explanandum E under a law L, so that E can be derived from L in conjunction with particular facts. He also recognizes a distinctive kind of "inductive-statistical" (IS) explanation, in which E is subsumed under a *statistical* law, which will take the form of a statement of conditional probability; in this case, E cannot be validly derived from the law and particular facts, but rather is rendered probable in accordance with the conditional probability.

Confirmation

While 'correlation is causation' is an unpromising slogan, 'correlation is confirmation' has fared much better. Confirmation is a useful concept, because even if Hume was right that there are no necessary connections between distinct existences, still it seems there are at least some non-trivial probabilistic relations between them. That's just what we mean by saying things like 'B supports A', or 'B is evidence for A', or 'B is counterevidence for A', or 'B disconfirms A'. So, many Bayesians appropriate the unsuccessful first attempt above to analyze causation, and turn it into a far more successful attempt to analyze confirmation — confirmation is positive correlation, disconfirmation is negative correlation, and evidential irrelevance is independence:

Relative to probability function P,

- E confirms H iff $P(H|E) > P(H)$
- E disconfirms H iff $P(H|E) < P(H)$
- E is evidentially irrelevant to H iff $P(H|E) = P(H)$

Curve-fitting, and the Akaike Information Criterion

Scientists are familiar with the problem of fitting a curve to a set of data. Forster and Sober [1994] argue that the real problem is one of trading off verisimilitude and simplicity: for a given set of data points, finding the curve that best balances the desiderata of predicting the points as accurately as possible using a function that has as few parameters as possible, so as not to 'overfit' the points. They argue that simplicity should be attributed to *families* of curves rather than to individual curves. They advocate selecting the family F with the best expected 'predictive accuracy', as measured by the Akaike Information Criterion:

$$AIC(F) = \frac{1}{N}\log(P(\text{Data}|L(F)) - k),$$

where $L(F)$ is the member of F that fits the data best, and k is the number of adjustable parameters of members of F.

Various other approaches to the curve-fitting problem similarly appeal to likelihoods (at least tacitly), and thus to conditional probabilities. They include the Bayesian Information Criterion [BIC; Schwarz, 1978], Minimum Message Length inference [MML; Wallace and Dowe, 1999; Dowe *et al.*, 2007] and Minimum Description Length inference [MDL; Grunwald *et al.*, 2005].

Decision theory

Decision theory purports to tell us how an agent's beliefs and desires in tandem determine what she should do. It combines her utility function and her probability function to give a figure of merit for each possible action, called the *expectation*, or *desirability* of that action (rather like the formula for the expectation of a random variable): a weighted average of the utilities associated with each action. In so-called 'evidential decision theory', as presented by Jeffrey [1983], the weights are conditional probabilities for states, given actions. Let $S_1, S_2, ..., S_n$ be a partition of possible states of the world. The choice-worthiness of action A is given by:

$$V(A) = \sum_i u(A \& S_i) P(S_i | A)$$

And so it goes again — this has just been another sampler. Given the obvious importance of conditional probability in philosophy, it will be worth investigating how secure are its foundations in (RATIO).

4 PROBLEMS WITH THE RATIO ANALYSIS OF CONDITIONAL PROBABILITY

So far we have looked at success stories for the usual understanding of conditional probability, given by (RATIO). We have seen that several different kinds of argument triangulate to it, and that it subserves a vast variety of applications of conditional probability — indeed, this latter fact itself provides a further pragmatic argument in favour of it. But we have not yet reached the end of the story. I turn now to four different kinds of problem for the ratio analysis, each mitigating the arguments in its favour.

4.1 Conditions with probability zero

$\frac{P(A \cap B)}{P(B)}$ is undefined when $P(B) = 0$; the ratio formula comes with the proviso that $P(B) > 0$. The proviso would be of little consequence if we could be assured that all probability-zero events of any interest are impossible, but as probability textbooks and even Kolmogorov himself caution us, this is not so. That is, we could arguably dismiss probability zero antecedents as 'don't cares' if we could be assured that all probability functions of any interest are *regular* — that is, they assign probability 0 only to the empty set. But this is not so. Worse, there are many cases of such conditional probabilities in which intuition delivers a clear verdict as to the correct answer, but (RATIO) delivers no verdict at all.

Firstly, in uncountable probability spaces one cannot avoid probability zero events that are possible — indeed, we are saddled with uncountably many of them.[2] Consider probability spaces with points taken from a continuum. Here is an example originating with Borel: A point is chosen at random from the surface of the earth (thought of as a perfect sphere); what is the probability that it lies in the Western hemisphere, given that it lies on the equator? 1/2, surely. Yet the probability of the condition is 0, since a uniform probability measure over a sphere must award probabilities to regions in proportion to their area, and the equator

[2] Here I continue to assume Kolmogorov's axiomatization, according to which probabilities are *real-valued.* Regularity may be achieved by the use of *infinitesimal probabilities* (see e.g. [Skyrms, 1980]); but see [Hájek, 2003] for some concerns.

has area 0. The ratio analysis thus cannot deliver the intuitively correct answer. Obviously there are uncountably many problem cases of this form for the sphere.

Another class of problem cases arises from the fact that the power set of a denumerable set is uncountable. For example, the set of all possible infinite sequences of tosses of a coin is uncountable (the sets of positive integers that could index the heads outcomes form the power set of the positive integers). Any particular sequence has probability zero (assuming that the trials are independent and identically distributed with intermediate probability for heads). Yet surely various corresponding conditional probabilities are defined — e.g., the probability that a fair coin lands heads on every toss, given that it lands heads on tosses 3, 4, 5, ..., is $^1\!/_4$.

More generally, the various classic 'almost sure' convergence results — the strong law of large numbers, the law of the iterated logarithm, the martingale convergence theorem, etc. — assert that certain convergences take place, not with certainty, but 'almost surely'. This is not merely coyness, since these convergences may fail to take place — again, genuine possibilities that receive probability 0, and interesting ones at that. The fair coin may land heads on every toss, and it would be no less fair for it.

Zero probability events also arise naturally in countable probability spaces if we impose certain symmetry constraints (such as the principle of indifference), and if we are prepared to settle for finite additivity. Following de Finetti [1990], imagine an infinite lottery whose outcomes are equiprobable. Each ticket has probability 0 of winning, although with probability 1 some ticket will win. Again, various conditional probabilities seem to be well-defined: for example, the probability that ticket 1 wins, given that either ticket 1, 2, or 3 wins, is surely 1/3.

The problem of zero-probability conditions is not simply an artefact of the mathematics of infinite (and arguably idealized) probability spaces. For even in finite probability spaces, various possible events may receive probability zero. This is most obvious for subjective probabilities, and in fact it happens as soon as an agent updates on some non-trivial information, thus ruling out the complement of that information — e.g., when you learn that the die landed with an odd number showing up, thus ruling out that it landed with an even number showing up. But still it seems that various conditional probabilities with probability-zero conditions can be well-defined — e.g, the probability that the die landed 2, *given* that it landed 2, is 1. Indeed, it seems that there are some contingent propositions that one is rationally *required* to assign probability 0 — e.g., 'I do not exist'. But various associated conditional probabilities may be well-defined nonetheless — e.g. the probability that I do not exist, *given* that I do not exist, is 1. Perhaps such cases are more controversial. If so, it matters little. As long as there is *some* case of a well-defined conditional probability with a probability-zero condition, then (RATIO) is refuted as an analysis of conditional probability. (It may nonetheless serve many of our purposes well enough as a *partial* — that is, incomplete — analysis. See [Hájek, 2003] for further discussion.)

4.2 *Conditions with unsharp probability*

A number of philosophers and statisticians eschew the usual assumption that probabilities are always real numbers, sharp to infinitely many decimal places. Instead, probabilities may for example be intervals, or convex sets, or sets of real numbers more generally. Such probabilities are given various names: "indeterminate" [Levi, 1974; 2000], "vague" [van Fraassen, 1990], "imprecise" [Walley, 1991], although these words have other philosophical associations that may not be intended here. Maybe it is best to mint a new word for this purpose. I will call them *unsharp*, marking the contrast to the usual sharp probabilities, while remaining neutral as to how unsharp probabilities should be modelled.

What is the probability that the Democrats win the next U.S. election? Plausibly, the answer is unsharp. This is perhaps clearest if the probability is subjective. If you say, for example, that your credence that they win is 0.6, it is doubtful that you really mean 0.60000 . . . , precise to infinitely many decimal places. Now, what is the probability that the Democrats win the next U.S. election, *given* that they win the next U.S. election? Here the answer is sharp: 1. Or what is the probability that this fair coin will land Heads when tossed, given the Democrats win the next U.S. election? Again, the answer seems to be sharp: $1/2$. In Hájek (2003) I argue that cases like these pose a challenge to the ratio analysis: it seems unable to yield such results. To be sure, perhaps that analysis coupled with suitable devices for handling unsharpness — e.g. supervaluation — can yield the results (although I argue that they risk being circular). Still, the point remains that the ratio analysis cannot be the complete story about conditional probability.

4.3 *Conditions with vague probability*

A superficially similar, but subtly different kind of case involves conditions with what I will call *vague* probability. Let us first be clear on the distinction between unsharpness and vagueness in general, before looking at probabilistic cases (this is a reason why I did not adopt the word "vagueness" in the previous section). The hallmark of vagueness is often thought to be the existence of *borderline* cases. A predicate is vague, we may say, if there are possible individuals that do not clearly belong either to the extension or the anti-extension of the predicate. For example, the predicate "fortyish" is vague, conveying a fuzzy region centered around 40 for which there are borderline cases (e.g. a person who is 43). By contrast, I will think of an unsharp predicate as admitting of a range of possible cases, but not borderline cases. "Forty-something" is unsharp: it covers the range of ages in the interval [40, 50), but any particular person either clearly falls under the predicate or clearly does not. However we characterize the distinction, the phenomena of vagueness and unsharpness appear to be different. I now turn to the problem that *vague* probability causes for the ratio analysis.

Suppose that we run a million-ticket lottery. What is the probability that a *large-numbered* ticket wins? It is vague what counts as a 'large number' — 17 surely doesn't, 999,996 surely does, but there are many numbers that are not so easily classified. The probability assignment plausibly inherits this vagueness — it might be, for example, '0.3-ish', again with borderline cases. Now, what is the probability that a large-numbered ticket wins, *given* that a large-numbered ticket wins? That is surely razor-sharp: 1. As before, the challenge to the ratio analysis is to do justice to these facts.

4.4 Conditions with undefined probability

Finally, we come to what I regard as the most important class of problem cases for (RATIO), for they are so widespread and often mundane. They arise when neither $P(A \cap B)$ nor $P(B)$ is defined, and yet the probability of A, given B, is defined. Here are two kinds of case, the first more intuitive, the second more mathematically rigorous, both taken from [Hájek, 2003].

The first involves a coin that you believe to be fair. What is the probability that it lands heads, given that I toss it fairly? 1/2, of course. According to the ratio analysis, it is

$$P(\text{the coin lands heads} \mid \text{I toss the coin fairly}), \text{ that is,}$$

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

However, these unconditional probabilities may not be defined — e.g. you may simply not assign them values. After some thought, you may start to assign them values, but the damage has already been done; and then again, you may still not do so. In [Hájek, 2003] I argue that this ratio may well remain undefined, and I rebut various proposals for how it may be defined after all.

The second kind of case involves non-measurable sets. Imagine choosing a point at random from the [0, 1] interval. We would like to model this with a uniform probability distribution, one that assigns the same probability to a given set as it does to any translation (modulo 1) of that set. Assuming the axiom of choice and countable additivity, it can be shown that for any such distribution P there must be sets that receive no probability assignment at all from P — so called 'non-measurable sets'. Let N be such a set. Then $P(N)$ is undefined. Nonetheless, it is plausible that the probability that the chosen point comes from N, given that it comes from N, is 1; the probability that it does not come from N, given that it comes from N, is 0; and so on. The ratio analysis cannot deliver these results.

The coin toss case may strike you as contentious, and the non-measurable case as pathological (although in [Hájek, 2003] I defend them against these charges). But notice that many of the paradigmatic applications of conditional probability canvassed in the previous section would seem to go the same way. For example, the Born rule surely should *not* be understood as assigning a value to a ratio of unconditional probabilities of the form

$$\frac{P(\text{measurement outcome } O_k \text{ is observed} \cap \text{measurement } M \text{ is performed})}{P(\text{measurement } M \text{ is performed})}.$$

Among other things, the terms in the ratio are clearly not given by quantum mechanics, and may plausibly not be defined at all, involving as they do a tacit quantification over the free actions of an open-ended set of experimenters.

To summarize: we have seen four kinds of case in which the ratio analysis appears to run aground: conditional probabilities with conditions whose probabilities are either zero, unsharp, vague, or undefined. Now there is a good sense in which these are problems with *unconditional* probability in its own right, which I am parlaying into problems for conditional probability. For example, the fact that Kolmogorov's theory of unconditional probability conflates zero-probability possibilities with genuine impossibilities may seem to be a defect of that theory, quite apart from its consequences for conditional probability. Still, since his theory of conditional probability is parasitic on his theory of unconditional probability, it should come as no surprise that defects in the latter can be exploited to reveal defects in the former. And notice how the problems in unconditional probability theory can be *amplified* when they become problems in conditional probability theory. For example, the conflation of zero-probability possibilities with genuine impossibilities might be thought of as a minor 'blurriness in vision' of probability theory; but it is rather more serious when it turns into problems of outright undefinedness in conditional probability, total blind spots.

Here are two ways that one might respond. First, one might preserve the conceptual priority that Kolmogorov gives to unconditional over conditional probability, but seek a more sophisticated account of conditional probability. Second, one might reverse the conceptual order, and regard conditional probability as the proper primitive of probability theory. The next two sections discuss versions of these responses, respectively.

5 KOLMOROGOV'S REFINEMENT: CONDITIONAL PROBABILITY AS A RANDOM VARIABLE

(This section is more advanced, and may be skipped by readers who are more interested in philosophical issues. Its exposition largely follows [Billingsley, 1995]; the ensuing critical discussion is my own.)

Kolmogorov went on to give a more sophisticated account of conditional probability as a random variable.

5.1 Exposition

Let the probability space $\langle \Omega, \mathcal{F}, P \rangle$ be given. We will interpret P as the credence function of an agent, which assumes the value $P(\omega)$ at each point $\omega \in \Omega$. Fixing $A \in \mathcal{F}$, we may define the random variable whose value is:

$$P(A|B) \text{ if } \omega \in B,$$
$$P(A|B^c) \text{ if } \omega \in B^c.$$

Think of our agent as about to learn the result of the experiment regarding B, and she will update accordingly. (§7 discusses updating rules in greater detail.)

Now generalize from the 2-celled partition $\{B, B^c\}$ to any countable partition $\{B_1, B_2, \ldots\}$ of Ω into $\mathcal{F}$-sets. Let $\mathcal{G}$ consist of all of the unions of the B_i; it is the smallest sigma field that contains all of the B_i. $\mathcal{G}$ can be thought of as an experiment. Our agent will learn which of the B_i obtains — that is, the outcome of the experiment — and is poised to update her beliefs accordingly. Fixing $A \in \mathcal{F}$, consider the function whose values are:

$$\begin{aligned}&P(A|B_1) \text{ if } \omega \in B_1,\\ &P(A|B_2) \text{ if } \omega \in B_2,\\ &\ldots\end{aligned}$$

when these quantities are defined. If $P(B_i) = 0$, let the corresponding value of the function be chosen arbitrarily from $[0, 1]$, this value constant for all $\omega \in B_i$. Call this function the *conditional probability of A given* $\mathcal{G}$, and denote it $P[A\|\mathcal{G}]$. Given the latitude in assigning a value to this function if $P(B_i) = 0, P[A\|\mathcal{G}]$ stands for any one of a family of functions on Ω, differing on how this arbitrary choice is made. A specific such function is called a *version* of the conditional probability. Thus, any two versions agree except on a set of probability 0. Any version codifies all of the agent's updating dispositions in response to all of the possible results of the experiment.

Notice that since any $G \in \mathcal{G}$ is a disjoint union $\cup_k Bi_k$, the probability of any set of the form $A \cap G$ can be calculated by the law of total probability:

$$P(A \cap G) = \sum\nolimits_k P(A|Bi_k)P(Bi_k) \qquad (1)$$

We may generalize further to the case where the sigma field G may not necessarily come from a countable partition, as was previously the case. Our agent will learn for each G in $\mathcal{G}$ whether $\omega \in G$ or $\omega \in G^c$. Generalizing (1), we would like to be assured of the existence of a function $P[A\|\mathcal{G}]$ that satisfies the equation:

$$P(A \cap G) = \int_G P[A\|\mathcal{G}]dP \text{ for all } G \in \mathcal{G}.$$

That assurance is provided by the *Radon-Nikodym theorem*, which for probability measures ν and P defined on $\mathcal{F}$ states:

If $P(X) = 0$ implies $\nu(X) = 0$ then there exists a function f such that

$$\nu(A) = \int_A f dP$$

for all $A \in \mathcal{F}$.

Let $\nu(G) = P(A \cap G)$ for all $G \in \mathcal{G}$. Notice that $P(G) = 0$ implies $\nu(G) = 0$ so the Radon-Nikodym theorem applies: the function $P[A\|\mathcal{G}]$ that we sought does indeed exist. As before, there may be many such functions, differing on their assignments to probability-zero sets; any such function is called a *version* of the conditional probability.

Stepping back for a moment: $\int_G P[A\|\mathcal{G}]dP$ is the expectation of the random variable $P[A\|\mathcal{G}]$, conditional on G, weighted according to the measure P. We have come back full circle to the remark made earlier about the law of total probability: an unconditional probability can be identified with an expectation of probabilities conditional on each cell of a partition, weighted according to the unconditional probabilities of the cells.

5.2 Critical discussion

Kolmogorov's more sophisticated formulation of conditional probability provides some relief from the problem of conditions with probability zero — there is no longer any obstacle to such conditional probabilities being defined. However, the other three problems for the ratio analysis — conditions with unsharp, vague, or undefined probability — would appear to remain. For the more sophisticated formulation equates a certain integral, in which the relevant conditional probability figures, to the probability of a conjunction; but when this latter probability is either unsharp, vague, or undefined, the analysis goes silent.

Moreover, there is further trouble that had no analogue for the ratio analysis, as shown by Seidenfeld, Schervish, and Kadane in their [2001] paper on "regular conditional distributions" — i.e. distributions of the form $P[\_\|\mathcal{A}]$ that we have been discussing. Let $P[\_\|\mathcal{A}](\omega)$ denote the regular conditional distribution for the probability space $(\Omega, \mathcal{B}, P)$ given the conditioning sub-σ-field $\mathcal{A}$, evaluated at the point ω. Following Blackwell and Dubins [1975], say that a regular conditional distribution is *proper at* ω if it is the case that whenever $\omega \in A \in \mathcal{A}$,

$$P(A\|\mathcal{A})(\omega) = 1$$

The distribution is *improper* if it is not everywhere proper. Impropriety seems to be disastrous. We may hold this truth to be self-evident: the conditional probability of anything consistent, given *itself*, should be 1. Indeed, it seems to be about as fundamental fact about conditional probability as there could be, on a par with the fundamental fact in logic that any proposition implies itself. So the possibility of impropriety, however minimal and however localized it might be, is a serious defect in an account of conditional probability. But Seidenfeld *et al.* show just how striking the problem is. They give examples of regular conditional distributions that are *maximally* improper. They are cases in which

$$P[A\|\mathcal{A}](\omega) = 0$$

(as far from the desired value of 1 as can be), and this impropriety holds *almost everywhere* according to P, so the impropriety is maximal both locally and globally.[3] This is surely bad news for the more sophisticated analysis of conditional probability — arguably fatal.

6 CONDITIONAL PROBABILITY AS PRIMITIVE

A rival approach takes conditional probability $P(\_, \_)$ as primitive. If we like, we may then define the unconditional probability of a as $P(a, \mathbf{T})$, where $\mathbf{T}$ is a logical truth. (We use lower case letters and a comma separating them in keeping with Popper's formulation, which we will soon be presenting.) Various axiomatizations of primitive conditional probability have been defended in the literature. See Roeper and Leblanc [1999] for an encyclopedic discussion of competing theories of conditional probability, and Keynes [1921], Carnap [1950], Popper [1959b], and Hájek [2003] for arguments that probability is inherently a two-place function. As is so often the case, their work was foreshadowed by Jeffreys [1939/1961], who axiomatized a comparative conditional probability relation: p is more probable than q, given r.

In some ways, the most general of the proposed axiomatizations is Popper's [1959b], and his system is the one most familiar to philosophers. Renyi's [1970] axiomatization is undeservedly neglected by philosophers. It closely mimics Kolmogorov's axiomatization, replacing unconditional with conditional probabilities in natural ways. I regard it as rather more intuitive than Popper's system. But since the latter has the philosophical limelight, I will concentrate on it here.

Popper's primitives are: (i) Ω, the universal set; (ii) a binary numerical function $p(\_, \_)$ of the elements of Ω; a binary operation ab defined for each pair (a, b) of elements of Ω; a unary operation $\neg a$ defined for each element a of Ω. Each of these concepts is introduced by a postulate (although the first actually plays no role in his theory):

Postulate 1. The number of elements in Ω is countable.

Postulate 2. If a and b are in Ω, then $p(a, b)$ is a real number, and the following axioms hold:

A1. (Existence) There are elements c and d in Ω such that $p(a, b) \neq p(c, d)$.

A2. (Substitutivity) If $p(a, c) = p(b, c)$ for every c in Ω, then $p(d, a) = p(d, b)$ for every d in Ω.

A3. (Reflexivity) $p(a, a) = p(b, b)$.

Postulate 3. If a and b are in Ω, then ab is in Ω; and if c is also in Ω, then the following axioms hold:

[3] A necessary condition for this is that the conditioning sub-sigma algebra is not countably generated.

B2. (Monotony) $p(ab, c) \leq p(a, c)$

B2. (Multiplication) $p(ab, c) = p(a, bc)p(b, c)$

Postulate 4. If a is in Ω, then $\neg a$ is in Ω; and if b is also in Ω, then the following axiom holds:

C. (Complementation) $p(a, b) + p(\neg a, b) = p(b, b)$, unless $p(b, b) = p(c, b)$ for every c in Ω.

Popper also adds a "fifth postulate", which may be thought of as giving the definition of absolute (unconditional) probability:

Postulate AP. If a and b are in Ω, and if $p(b, c) \geq p(c, b)$ for every c in Ω, then $p(a) = p(a, b)$.

Popper's axiomatization thus generalizes ordinary probability theory. Intuitively, b can be regarded as a logical truth. Unconditional probability, then, can be regarded as probability conditional on a logical truth. However, a striking fact about the axiomatization is that it is *autonomous* — it does not presuppose any set-theoretic or logical notions (such as "logical truth"). A function $p(\_, \_)$ that satisfies the above axioms is called a *Popper function*.

A well-known advantage of the Popper function approach is that it allows conditional probabilities of the form $p(a, b)$ to be defined, and to have intuitively correct values, even when the 'condition' b has absolute probability 0, thus rendering the usual conditional probability ratio formula inapplicable — we saw examples in §4.1. Moreover, Popper functions can bypass our concerns about conditions with unsharp, vague, or undefined probabilities — the conditional probabilities at issue are assigned directly, without any detour or constraint given by unconditional probabilities.

McGee [1994] shows that in an important sense, probability statements cast in terms of Popper functions and those cast in terms of nonstandard probability functions are inter-translatable. If r is a nonstandard real number, let *st(r)* denote the *standard part of* r, that is, the unique real number that is infinitesimally close to r. McGee proves the following theorem: If P is a nonstandard-valued probability assignment on a language $\mathcal{L}$ for the classical sentential calculus, then the function $C : \mathcal{L} \times \mathcal{L} \rightarrow \mathbf{R}$ given by

$$\begin{aligned} C(a, b) &= st(\frac{P(ab)}{P(b)}), \text{ provided } P(b) > 0 \\ &= 1, \text{ otherwise} \end{aligned}$$

is a Popper function. Conversely, if C is a Popper function, there is a nonstandard-valued probability assignment P such that

$$P(b) = 1 \text{ iff } C(\_, b) \text{ is the constant function } 1$$

and

$$C(c,b) = st(\frac{P(cb)}{P(b)}) \text{ whenever } P(b) > 0.$$

The arguments adduced in §4 against the ratio analysis of conditional probability indirectly support taking conditional probability as primitive, although they also leave open the viability of some *other* analysis of conditional probability in terms of unconditional probability. However, there are some considerations that seem to favour the primacy of conditional probability.

The conditional probability assignments that I gave in §4's examples are seemingly non-negotiable. They can, and in some cases must, stand without support from corresponding unconditional probabilities. Moreover, the examples of unsharp, vague, and undefined probabilities suggest that the problem with the ratio analysis is not so much that it is a *ratio* analysis, but rather that it is a ratio *analysis*. 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 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 it can be regarded as conditional on tacit assumptions about the coin, the toss, the immediate environment, and so on. In fact, it can be regarded as conditional on your total evidence — recall the quotation from de Finetti in the second paragraph of this article. Now, perhaps in very special cases we can assign a probability free of all assumptions — an assignment of 1 to 'I exist' may be such a case. But even then, the probability is easily recovered as probability conditional on a logical truth or some other *a priori* truth. Furthermore, we can be sure that there can be no analogue of the 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 of X given a logical/*a priori* truth. Unconditional probabilities are special cases of conditional probabilities.

These considerations are supported further by our discussion in §3.1 of how according to the leading interpretations probability statements are always at least tacitly relativised — on the frequency interpretations, to a reference class; on the propensity interpretation, to a chance set-up; on the classical and logical interpretation, to a body of evidence; on the subjective interpretation, to a subject (who has certain background knowledge) at a time, and who may defer to some 'expert' (a person, a future self, a relative frequency, a chance).

Putting these facts together, we have a case for regarding 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. But I also recommend Kenny Easwaran's contribution to this volume ("The Varieties of Conditional Probability") for a different perspective.

7 CONDITIONAL PROBABILITIES AND UPDATING RULES

7.1 Conditionalization

Suppose that your degrees of belief are initially represented by a probability function $P_{initial}(\_)$, and that you become certain of E (where E is the strongest such proposition). What should be your new probability function P_{new}? The favoured updating rule among Bayesians is *conditionalization;* P_{new} is related to $P_{initial}$ as follows:

(Conditionalization) $P_{new}(X) = P_{initial}(X|E)$ (provided $P_{initial}(E) > 0$)

Conditionalization is supported by some arguments similar to those that supported the ratio analysis. Firstly, there is case-by-case evidence. Upon receiving the information that the die landed odd, intuition seems to judge that your probability that it landed 5 should be revised to 1/3, just as conditionalization would have it. Similarly for countless other judgments. Secondly, the muddy Venn diagram can now be given a dynamic interpretation: learning that E corresponds to scraping all mud off $\neg E$. What to do with the mud that remains? It obviously must be rescaled, since it amounts to a total of only $P_{initial}(E)$, whereas probabilities must sum to 1. Moreover, since nothing stronger than E was learned, any movements of mud within E seem gratuitous, or even downright unjustified. So our desired updating rule should preserve the profile of mud within E but renormalize it by a factor of $1/P_{initial}(E)$; this is conditionalization.

Thirdly, conditionalization is supported by a 'diachronic' Dutch Book argument (see [Lewis, 1999]): on the assumption that your updating is rule-governed, you are subject to a Dutch book (with bets placed at different times) if you do not conditionalize. Equally important is the converse theorem [Skyrms, 1987]: if you do conditionalize, then you are immune to such a Dutch Book.

Then there are arguments for conditionalization for which there are currently no analogous arguments for (RATIO) — although I suggest that it would be fruitful to pursue such arguments. For example, Greaves and Wallace [2006] offer a "cognitive decision theory", arguing that conditionalization is the unique updating rule that maximizes expected *epistemic* utility.

However, there are also some sources of suspicion and even downright dissatisfaction about conditionalization. There are apparently some kinds of belief revisions that should *not* be so modelled. Those involving *indexical* beliefs are a prime example. I am currently certain that my computer's clock reads 8:33; and yet by the time I reach the end of this sentence, I find that I am certain that it does *not* read 8:33. Probability mud isn't so much scraped away as *pushed sideways* in such cases. Levi [1980] insists that conditionalization is also not appropriate in cases where an agent "contracts" her "corpus" of beliefs — when her stock of settled assumptions is somehow challenged, forcing her to reduce it. See [Hild, 1998;

Bacchus *et al.*, 1990; Arntzenius, 2003] for further objections to conditionalization.

Much as the considerations supporting conditionalization are similar to those supporting the ratio analysis, the considerations *counter*-supporting the latter *counter*-support the former. In particular, the objections that I raised in §4 would seem to have force equally against the adequacy of conditionalization. Recall the problem of conditions with probability zero. A point has just been chosen at random from the surface of the earth, and you learn that it lies on the equator. Conditionalization cannot model this revision in your belief state, since you previously gave probability zero to what you learned. (The same would be true of any line of latitude on which you might learn the point to be.) Similarly for your learning that Democrats won the U.S. election; similarly for your learning that a large-numbered ticket was picked in the 1,000,000-ticket lottery; similarly for your learning that I tossed the fair coin; similarly for your learning that a randomly chosen point came from the non-measurable set N.

To be sure, the key idea behind conditionalization can be upheld while disavowing the ratio analysis for conditional probability. Upon receiving evidence E, one's new probability for X should be one's initial conditional probability for X, given E — this is neutral regarding how the conditional probability should be understood. My point is that the standard formulation of conditionalization, stated above, is *not* neutral: it presupposes the ratio analysis of conditional probability and inherits its problems. (Recall that $P(\_|\_)$ is shorthand for a ratio of unconditional probabilities.)

Popper functions allow a natural reformulation of updating by conditionalization, so that even items of evidence that were originally assigned such problematic unconditional probabilities by an agent can be learned. The result of conditionalizing a Popper function $P(\_,\_)$ on a piece of evidence encapsulated by e is $P(\_,\_e)$ — for example, $P(a,b)$ gets transformed to $P(a,be)$.

7.2 Jeffrey conditionalization

Jeffrey conditionalization allows for less decisive learning experiences in which your probabilities across a partition $\{E_1, E_2, ...\}$ change to $\{P_{new}(E_1), P_{new}(E_2), ...\}$, where none of these values need be 0 or 1:

$$P_{new}(X) = \sum_i P_{initial}(X|E_i)P_{new}(E_i)$$

[Jeffrey, 1965; 1983; 1990]. Notice that if we replace $P_{initial}(X|E_i)$ by $P_{new}(X|E_i)$, we simply have an instance of the law of total probability. This theorem of the probability calculus becomes a norm of belief revision, assuming that probabilities conditional on each cell of the partition should stay 'rigid', unchanged throughout such an experience. Diaconis and Zabell [1982] show, by reasonable criteria for determining a metric on the space of probability functions, that this rule corresponds to updating to the nearest function in that space, subject to the constraints. One might interpret this as capturing a kind of epistemic conservatism in the spirit of a Quinean “minimal mutilation” principle: staying as ‘close’ to your original opinions as you can, while respecting your evidence.

Jeffrey conditionalization is again supported by a diachronic Dutch book argument [Armendt, 1980]. It should be noted, however, that diachronic Dutch Book arguments have found less favour than their synchronic counterparts. Levi [1991] and Maher [1992] insist that the agent who fails to conditionalize and who thereby appears to be susceptible to a Dutch Book will be able to ‘see it coming’, and thus avoid it; however, see also Skyrms’ [1993] rebuttal. Christensen [1991] denies that the alleged ‘inconsistency’ dramatized in such arguments has any normative force in the diachronic setting. van Fraassen [1989] denies that rationality requires one to follow a rule in the first place. Levi [1967] also criticizes Jeffrey conditionalization directly. For example, repeated operations of the rule may not commute, resulting in a path-dependence of one’s final epistemic state that might be found objectionable. However, Lange [2000] argues that this non-commutativity is a virtue rather than a vice.

8 SOME PARADOXES AND PUZZLES INVOLVING CONDITIONAL PROBABILITY AND CONDITIONALIZATION

8.1 The Monty Hall problem

Let’s begin with a problem that is surely *not* a paradox, even though it is often called that. You are on the game show *Let’s Make a Deal* hosted by Monty Hall. Before you are three doors; behind exactly one of them is a prize, which you will win if you choose its door correctly. First, you are to nominate a door. Monty, who knows where the prize is and will not reveal it, ostentatiously opens *another* door, revealing it to be empty. He then gives you the opportunity to switch to the remaining door. Should you do so? Many people intuit that it doesn’t matter either way: you’re as likely to win the prize by sticking with your original door as you are by switching. That’s wrong — indeed, you are twice as likely to win by switching than by sticking with your original door. An easy way to see this is to consider the probability of *failing* to win by switching. The only way you could fail would be if you had initially nominated the correct door — probability 1/3 — and then, unluckily, switched away from it when given the chance. Thus, the probability of winning by switching is 2/3.

The reasoning just given is surely too simple to count as paradoxical. But the problem does teach a salutary lesson regarding the importance on conditionalizing on one’s *total* evidence. The fallacious reasoning would have you conditionalize on the evidence that the prize is not behind the door that Monty actually opens (e.g. door 1) — that is, to assign a probability $1/2$ to each of the two remaining doors (e.g. doors 2 and 3). But your actual evidence was stronger than that: you also learned that Monty *opened the door that he did*. (If you initially chose the correct door, he had a genuine choice.) A relatively simple calculation shows that conditionalizing on your total evidence yields the correct answer: your updated probability that the remaining door contains the prize is 2/3, so you should switch to it.

8.2 Simpson's paradox

Again, it is questionable whether an observation due to Simpson deserves to be called a "paradox"; rather, it is a fairly straightforward fact about inequalities among conditional probabilities. But the observation is undoubtedly rather counterintuitive, and it has some significant ramifications for scientific inference.

The paradox was once famously instantiated by the U.C. Berkeley's admission statistics. Taken as a whole, admissions seemed to favour males, as suggested by the correlations inferred from the relative frequencies of admission of males and females:

$P(\text{admission} \mid \text{male}) > P(\text{admission} \mid \text{female})$.

Yet disaggregating the applications department by department, the correlations went the other way:

$P(\text{admission} \mid \text{male \& department 1 applicant}) < P(\text{admission} \mid \text{female \& department 1 applicant})$

$P(\text{admission} \mid \text{male \& department 2 applicant}) < P(\text{admission} \mid \text{female \& department 2 applicant})$,

and so on for every department.

How could this be? A simple explanation was that the females tended to apply to more competitive departments with lower admission rates. This lowered their university-wide admission rate compared to males, even though department by department their admission rate was superior.

More generally, Simpson's paradox is the phenomenon that correlations that appear at one level of partitioning may disappear or even reverse at another level of partitioning:

$P(E|C) > P(E|\sim C)$ is consistent with

$P(E|C \;\&\; F_1) < P(E|\sim C\& \; F_1)$,

$P(E|C \;\&\; F_2) < P(E|\sim C \;\&\; F_2)$,

$\ldots$

$P(E|C \;\&\; F_n) < P(E|\sim C \;\&\; F_n)$,

for some partition $\{F_1, F_2, \ldots, F_n\}$.

Pearl [2000] argues that such a pattern of inequalities only seems paradoxical if we impose a causal interpretation on them. In our example, being male is presumably regarded as a (probabilistic) *cause* of being admitted, perhaps due to discrimination in favour of men and against women. We seem to be reasoning: "Surely *unanimity* in the departmental causal facts has to be preserved by the university at large!" Pearl believes that if we rid ourselves of faulty intuitions about correlations revealing causal relations, the seeming paradoxicality will vanish.

I demur. I think that we are just as liable to recoil even if the data is presented as inequalities among ratios, with no causal interpretation whatsoever. Department by department, the *ratio* of admitted women is greater than the *ratio* of admitted men, yet university-wide the inequality among the *ratios* goes the other way. How could this be? "Surely *unanimity* in the departmental ratio inequalities has to be preserved by the university at large!" Not at all, as simple arithmetic proves. We simply have faulty arithmetical intuitions.

8.3 The Judy Benjamin problem

The general problem for probability kinematics is: given a prior probability function P, and the imposition of some constraint on the posterior probability function, what should this posterior be? This problem apparently has a unique solution for certain constraints, as we have seen — for example:

1. Assign probability 1 to some proposition E, while preserving the relative odds of all propositions that imply E. Solution: conditionalize P on E.

2. Assign probabilities $p_1, ..., p_n$ to the cells of the partition $\{E_1, ..., E_n\}$, while preserving the relative odds of all propositions within each cell. Solution: Jeffrey conditionalize P on this partition, according to the specification.

But consider the constraint:

3. Assign conditional probability p to B, given A.

The *Judy Benjamin problem* is that of finding a rule for transforming a prior, subject to this third constraint [van Fraassen, 1989].

van Fraassen provides arguments for three distinct such rules, and surmises that this raises the possibility that such uniqueness results "will not extend to more broadly applicable rules in general probability kinematics. In that case rationality will not dictate epistemic procedure even when we decide that it shall be rule governed" [1989, p. 343].

8.4 Non-conglomerability

Call P *conglomerable in the partition* $X = \{x_1, x_2, \ldots\}$ if

$$k_1 \leq P(Y) \leq k_2 \text{ whenever } k_1 \leq P(Y|X = x_i) \leq k_2 \text{ for all } i = 1, 2, \ldots$$

Here's the intuitive idea. Suppose that you know now that you will learn which member of a particular partition is true. (A non-trivial partition might have as few as two members, such as {Heads} and {Tails}, or it might have countably many members.) Suppose further that you know now that *whatever* you learn, your probability for Q will lie in a certain interval. Then it seems that you should now assign a probability for Q that lies in that interval. If you *know* that you are going to have a certain opinion in the future, why wait? — Make it your opinion *now!* More generally, if you *know* that a credence of yours will be bounded in a particular way in the future, why wait? — Bound that credence in this way *now!* 'Conglomerability in a partition' captures this desideratum.

Failures of conglomerability arise when P is finitely additive, but not countably additive. As Seidenfeld *et al.* [1998] show, in that case there exists some countable partition in which P is not conglomerable. If updating takes place by conditionalization, failures of conglomerability lead to curious commitments reminiscent of violations of the Reflection Principle: "My future self, who is ideally rational and better informed than I am, will definitely have a credence for Q in a particular interval, but my credence for Q is not in this interval." (See [Jaynes, 2003, Ch. 15] for a critique of Seidenfeld *et al.* See also Kadane *et al.* [1986] for a non-conglomerability result even assuming countable additivity, in uncountable partitions.)

8.5 The two-envelope paradoxes

As an example of non-conglomerability, consider the following infinite version of the so-called 'two envelope' paradox: Two positive integers are selected at random and turned into dollar amounts, the first placed in one envelope, the second placed in another, whose contents are concealed from you. You get to choose one envelope, and its contents are yours. Suppose that following de Finetti [1972], and in violation of countable additivity, you assign probability 0 to all finite sets of positive integers, but (of course), probability 1 to the entire set of positive integers. Let X be the amount in your envelope and Y the amount in the other envelope. Then very reasonably you assign:

$$P(X < Y) = 1/2.$$

But suppose now that we let you open your envelope. You may see \$1, or \$2, or \$3, or ... Yet whatever you see, you will want to switch to holding the other envelope, for

$$P(X < Y|X = x) = 1 \text{ for } x = 1, 2, \ldots$$

Why wait? Since you know that you will want to switch, you should switch now. That is absurd: you surely cannot settle from the armchair that you have made the wrong choice, however you choose.

A better-known version of the two-envelope paradox runs as follows. One positive integer is selected and that number of dollars is placed in an envelope. Twice as much is placed in another envelope. The contents of both envelopes are concealed from you. You get to choose one envelope, and its contents are yours. At first you think that you have no reason to prefer one envelope over another, so you choose one. But as soon as you do, you feel regret. You reason as follows: "I am holding some dollar amount — call it n. The other envelope contains either

$2n$ or $n/2$, each with probability $1/2$. So its expectation is $(2n)1/2 + (n/2)1/2 = 5n/4 > n$. So it is preferable to my envelope." This is already absurd, as before. Worse, if we let you switch, your regret will immediately run the other way: "I am holding some dollar amount – call it m ... " And similar reasoning seems to go through even if we let you open your envelope to check its contents!

Let X be the random variable 'the amount in your envelope', and let Y be 'the amount in the other envelope'. Notice that a key step of the reasoning moves from

$$\text{for any } n, E(Y|X = n) > n = E(X|X = n) \qquad (*)$$

to the conclusion that the other envelope is preferable. A missing premise is that

$$E(Y) > E(X).$$

This may seem to follow straightforwardly from $(*)$. But that presupposes conglomerability with respect to the partition of amounts in your envelope, which is exactly what should be questioned. See [Arntzenius and McCarthy, 1997; Chalmers, 2002] for further discussion.

9 PROBABILITIES OF CONDITIONALS AND CONDITIONAL PROBABILITIES

A number of authors have proposed that there are deep connections between conditional probabilities and conditionals. Ordinary English seems to allow us to shift effortlessly between the two kinds of locutions. 'The probability of it raining, given that it is cloudy, is high' seems to say the same thing as 'the probability of it raining, if it is cloudy, is high' — the former a conditional probability, the latter the probability of a conditional.

The Ramsey test and Adams' thesis

Ramsey [1931/1990, p. 155] apparently generalized this observation in a pregnant remark in a footnote: "If two people are arguing 'If p will q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; ... We can say they are fixing their degrees of belief in q given p." Adams [1975] more explicitly generalized the observation in his celebrated thesis that the probability of the indicative conditional 'if A, then B' is given by the corresponding conditional probability of B given A. He denied that such conditionals have truth conditions, so this probability is not to be thought of as the probability that 'if A, then B' is true. Further, Adams' 'probabilities' of conditionals do not conform to the usual probability calculus — in particular, Boolean compounds involving them do not receive 'probabilities', as the usual closure assumptions (given in §2.1) would require.

For this reason, Lewis [1976] suggests that they be called "assertabilities" instead, a practice that has been widely adopted subsequently. Note, however, that "assertability" seems to bring in the norms of *assertion.* For example, Williamson [2002] argues that you should only assert what you *know*; but then it is hard to make sense of assertability coming in all the degrees that Adams requires of it. And conditionals can be unassertable for all sorts of reasons that seem beside the point here — they can be inappropriate, irrelevant, uninformative, undiplomatic, and so on. This is a matter of the pragmatics of conversation, which is another topic. Perhaps the locution "degrees of acceptability" better captures Adams' idea.

Stalnaker's Hypothesis

Stalnaker [1970], by contrast, insisted that conditionals have truth conditions, and he and Lewis were engaged in the late 60s and early 70s in a famous debate over what they were. In particular, they differed over the status of *conditional excluded middle* — on whether sentences of the following form are tautologies or not:

(CEM) $$(A \to B) \vee (A \to \neg B)$$

Stalnaker thought so; Lewis thought not. Stalnaker upheld the equality of genuine *probabilities* of conditionals with the corresponding conditional probabilities, and used the attractiveness of this thesis as an argument for his preferred semantics. More precisely, the hypothesis is that some suitably quantified and qualified version of the following equation holds:

(PCCP) $P(A \to B) = P(B|A)$ for all A, B in the domain of P, with $P(A) > 0$.

("$\to$" is a conditional connective.)

Stalnaker's guiding idea was that a suitable version of the hypothesis would serve as a criterion of adequacy for a truth-conditional account of the conditional. He explored the conditions under which it would be reasonable for a rational agent, with subjective probability function P, to believe a conditional $A \to B$. By identifying the probability of $A \to B$ with $P(B|A)$, Stalnaker was able to put constraints on the truth conditions of the '$\to$'. In particular, if this identification were sound, it would vindicate conditional excluded middle. For by the probability calculus,

$$\begin{aligned} & P[(A \to B) \vee (A \to \neg B)] \\ = \; & P(A \to B) + P(A \to \neg B) \\ & \text{(assuming that the disjuncts are incompatible, as both authors did)} \\ = \; & P(B|A) + P(\neg B|A) \\ & \text{(by the identification of probabilities of conditionals with conditional} \\ & \text{probabilities)} \\ = \; & 1. \end{aligned}$$

So all sentences of the CEM form have probability 1, as Stalnaker required.

Some of the probabilities-of-conditionals literature is rather unclear on exactly what claims are under discussion: what the relevant quantifiers are, and their domains of quantification. With the above motivations kept in mind, and for their independent interest, we now consider four salient ways of rendering precise the hypothesis that probabilities of conditionals are conditional probabilities:

Universal version: There is some $\rightarrow$ such that for all P, (PCCP) holds.

Rational Probability Function version: There is some $\rightarrow$ such that for all P that could represent a rational agent's system of beliefs, (PCCP) holds.

Universal Tailoring version: For each P there is some $\rightarrow$ such that (PCCP) holds.

Rational Probability Function tailoring version: For each P that could represent a rational agent's system of beliefs, there is some $\rightarrow$ such that (PCCP) holds.'indexorthogonal

Can any of these versions be sustained? The situation is interesting however we answer this question. If the answer is 'no', then seemingly synonymous locutions are not in fact synonymous: surprisingly, 'the probability of B, given A' does not mean the same thing as 'the probability of: B if A'. If the answer is 'yes', then important links between logic and probability theory will have been established, just as Stalnaker and Adams hoped. Probability theory would be a source of insight into the formal structure of conditionals. And probability theory in turn would be enriched, since we could characterize more fully what the usual conditional probability ratio means, and what its use is. de Finetti [1972] laments that (RATIO) gives the formula, but not the *meaning*, of conditional probability. A suitably quantified hypothesis involving (PCCP) could serve to characterize more fully what the ratio means, and what its use is.

There is now a host of results — mostly negative — concerning PCCP. We will give a sample of some of the most important ones. We will then be in a position to assess how the four versions of the hypothesis fare, and what the prospects are for other versions. Some preliminary definitions will assist in stating the results.

If (PCCP) holds, we will say that $\rightarrow$ is a *PCCP-conditional for P*, and that P is a *PCCP-function for $\rightarrow$*. If (PCCP) holds for a particular $\rightarrow$ for each member P of a class of probability functions $\mathcal{P}$, we will say that $\rightarrow$ is a *PCCP-conditional for $\mathcal{P}$*. A pair of probability functions P and P' are *orthogonal* if, for some $A, P(A) = 1$ but $P'(A) = 0$. (Intuitively, orthogonal probability functions concentrate their probability on entirely non-intersecting sets of propositions.) Call a proposition A a *P-atom* iff $P(A) > 0$ and, for all X, either $P(AX) = P(A)$ or $P(AX) = 0$. (Intuitively, a P-atom is a proposition that receives an indivisible 'blob' of probability from P.) Finally, we will call a probability function *trivial* if it has at most 4 different values.

Most of the negative results are 'triviality results': given certain assumptions, only trivial probability functions can sustain PCCP. Moreover, most of them make no assumptions about the logic of the '$\rightarrow$' — it is simply a two-place connective. The earliest and most famous results are due to Lewis [1976]:

> *First triviality result:* There is no PCCP-conditional for the class of all probability functions.

> *Second triviality result:* There is no PCCP-conditional for any class of probability functions closed under conditionalizing, unless the class consists entirely of trivial functions.

Lewis [1986] strengthens these results:

> *Third triviality result:* There is no PCCP-conditional for any class of probability functions closed under conditionalizing restricted to the propositions in a single finite partition, unless the class consists entirely of trivial functions.
>
> *Fourth triviality result:* There is no PCCP-conditional for any class of probability functions closed under Jeffrey conditionalizing, unless the class consists entirely of trivial functions.

These results refute the Universal version of the hypothesis. They also spell bad news for the Rational Probability Function version, for even if rationality does not require updating by conditionalizing, or Jeffrey conditionalizing, it seems plausible that it at least *permits* such updating. This version receives its death blow from the following result by Hall [1994], that significantly strengthens Lewis' results:

> *Orthogonality result:* Any two non-trivial PCCP-functions defined on the same algebra of propositions are orthogonal.

It follows from this that the Rational Probability Function version is true only if any two distinct rational agents' probability functions are orthogonal — which is absurd.

So far, the 'tailoring' versions remain unscathed. The Universal Tailoring version is refuted by the following result due to Hájek [1989; 1993], which concerns probability functions that assume only a finite number of distinct values:

> *Finite-ranged Functions Result:* Any non-trivial probability function with finite range has no PCCP-conditional.

This result also severely casts doubt on the Rational Probability Tailoring version, for it is hard to see why rationality requires one to adopt a probability function with infinite range. The key idea behind this result can be understood by considering a very simple case. Consider a three-ticket lottery, and let L_i = 'ticket i wins', $i = 1, 2, 3$. Let P assign probability $1/3$ to each of the L_i. Clearly, some conditional probabilities take the value $1/2$ — for example, $P(L_1|L_1 \vee L_2)$. But no unconditional probability can take this value, being constrained to be a multiple of $1/3$; *a fortiori,* no (unconditional) probability of a conditional can take this value. The point generalizes to all finite-ranged probability functions: there will always be some value of the conditional probability function that finds no match among the unconditional probabilities, and *a fortiori* no match among the (unconditional) probabilities of conditionals. Picture a dance at which, for a given finite-ranged probability function, all of the probability-of-a-conditional values line up along one wall, and all of the conditional probability values line up along the opposite wall.

Now picture each conditional probability value attempting to partner up with a probability of a conditional with the same value on the other side. According to Stalnaker's hypothesis, the dance would always be a complete success, with all the values finding their matches; the Finite-ranged Functions Result shows that the dance can never be a complete success. There will always be at least one wallflower among the conditional probabilities, which will have to sit out the dance — for example, $1/2$ in our lottery case.

If we make a minimal assumption about the logic of the $\rightarrow$, matters are still worse thanks to another result of Hall's [1994]:

> *No Atoms Result:* Let the probability space $\langle \Omega, \mathcal{F}, P \rangle$ be given, and suppose that PCCP holds for this P, and a '$\rightarrow$' that obeys modus ponens. Then $\langle \Omega, \mathcal{F}, P \rangle$ does not contain a P-atom, unless P is trivial.

It follows from this, on pain of triviality, that the range of P, and hence Ω and $\mathcal{F}$, are non-denumerable. All the more, it is hard to see how rationality requires this of an agent's probability space.

It seems, then, that all four versions of the hypothesis so far considered are untenable. (See also [Hájek, 1994] for more negative results.) For all that has been said so far, though, some suitably restricted 'tailoring' version might still survive. A natural question, then, is whether even Hall's 'no atoms' result can be extended — whether even uncountable probability spaces cannot support PCCP, thus effectively refuting any 'tailoring' version of the hypothesis. The answer is 'no' — and here we have a positive result due to van Fraassen [1976]. Suppose that $\rightarrow$ has this much logical structure:

(i) $[(A \rightarrow B) \cap (A \rightarrow C)] = [A \rightarrow (B \cap C)]$

(ii) $[(A \rightarrow B) \cup (A \rightarrow C)] = [A \rightarrow (B \cup C)]$

(iii) $[A \cap (A \rightarrow B)] = (A \cap B)$

(iv) $[A \rightarrow A] = \Omega$.

Such an $\rightarrow$ conforms to the logic *CE.* van Fraassen shows:

> *CE tenability result:* Any probability space can be extended to one for which PCCP holds, with an $\rightarrow$ that conforms to CE.

Of course, the larger space for which PCCP holds is uncountable. In the same paper, van Fraassen also shows that $\rightarrow$ can have still more logical structure, while supporting PCCP, provided we restrict the admissible iterations of $\rightarrow$ appropriately.

A similar strategy of restriction protects Adams' version of the hypothesis from the negative results. He applies PCCP to unembedded conditionals — 'simple' conditionals of the form $A \rightarrow B$, where A and B are themselves conditional-free. As mentioned before, Adams does not allow the assignment of probabilities to Boolean compounds involving conditionals; 'P' is thus not strictly speaking a probability function (and thus the negative results, which presuppose that it is, do not apply). McGee [1989] shows how Adams' theory can be extended to certain more complicated compounds of conditionals, while still falling short of full closure.

10 CONCLUSION

This survey has been long, and yet still I fear that some readers will be disappointed that I have not discussed adequately, or at all, their favourite application of or philosophical issue about conditional probability. They may find some solace in the lengthy bibliography that follows.

Along the way we have seen some reasons for questioning the orthodoxy enshrined in Kolmogorov's ratio analysis; moreover his more sophisticated formulation seems not entirely successful either. I have argued that we should take conditional probability as the primitive notion in probability theory, although this still remains a minority position. However we resolve this issue, we have something of a mathematical and philosophical balancing act: finding an entirely satisfactory mathematical and philosophical theory of conditional probability that as much as possible does justice to our intuitions and to its various applications. It is an act worth getting right: the foundations of probability theory depend on it, and thus any theory that employs probability theory depends on it also — which is to say, any serious empirical discipline, and much of philosophy.[4]

BIBLIOGRAPHY

[Adams, 1975] E. Adams. *The Logic of Conditionals,* Reidel, 1975.

[Armendt, 1980] B. Armendt. Is There a Dutch Book Argument for Probability Kinematics? *Philosophy of Science* 47, No. 4 (December), 583-88, 1980.

[Arntzenius, 2003] F. Arntzenius. Some Problems for Conditionalization and Reflection, *Journal of Philosophy* Vol. C, No 7, 356-371, 2003.

[Arntezenius and McCarthy, 1997] F. Arntzenius and D. McCarthy. The Two Envelope Paradox and Infinite Expectations, *Analysis* 57, No. 1, 28-34, 1997.

[Bacchus *et al.*, 1990] F. Bacchus, H. E. Kyburg Jr., and M. Thalos. Against Conditionalization, *Synthese* 85, 475-506, 1990.

[Billingsley, 1995] P. Billingsley. *Probability and Measure*, John Wiley and Sons, Third Edition, 1995.

[Blackwell and Dubins, 1975] D. Blackwell and L. Dubins. On Existence and Non-existence of Proper, Regular, Conditional Distributions, *The Annals of Probability* v3 i5, 741-752, 1975.

[Carnap, 1950] R. Carnap. *Logical Foundations of Probability*, Chicago: University of Chicago Press, 1950.

[Carnap, 1952] R. Carnap. *The Continuum of Inductive Methods*, Chicago: The University of Chicago Press, 1952.

[Cartwright, 1979] N. Cartwright. Causal Laws and Effective Strategies, *Noûs* **13**, 419-437, 1979.

[4]I am grateful to Elle Benjamin, Darren Bradley, Lina Eriksson, Marcus Hutter, Aidan Lyon, John Matthewson, Ralph Miles, Nico Silins, Michael Smithson, Weng Hong Tang, Peter Vanderschraaf, Wen Xuefeng, and especially John Cusbert, Kenny Easwaran and Michael Titelbaum for helpful suggestions.

[Chalmers, 2002] D. Chalmers. The St. Petersburg Two-Envelope Paradox, *Analysis* 62, 155-57, 2002.

[Christensen, 1991] D. Christensen. Clever Bookies and Coherent Beliefs, *The Philosophical Review* C, No. 2, 229-247, 1991.

[de Finetti, 1937] B. de Finetti. La Prévision: Ses Lois Logiques, Ses Sources Subjectives, *Annales de l'Institut Henri Poincaré*, 7: 1-68, 1937. Translated as Foresight. Its Logical Laws, Its Subjective Sources, in *Studies in Subjective Probability*, H. E. Kyburg, Jr. and H. E. Smokler (eds.), Robert E. Krieger Publishing Company, 1980.

[de Finetti, 1972] B. de Finetti. *Probability, Induction and Statistics*, Wiley, 1972.

[de Finetti, 1974/1990] B. de Finetti. *Theory of Probability*, Vol. 1. Chichester: Wiley Classics Library, John Wiley & Sons, 1974/1990.

[Diaconis and Zabell, 1982] P. Diaconis and S. L. Zabell. Updating Subjective Probability, *Journal of the American Statistical Association* 77, 822-30, 1982.

[Dowe *et al.*, 2007] D. Dowe, S. Gardner, and G. Oppy. Bayes not Bust! Why Simplicity is No Problem for Bayesians, *British Journal for Philosophy of Science* 58, 4, 709-54, 2007.

[Eells, 1991] E. Eells. *Probabilistic Causality,* Cambridge University Press, Cambridge, 1991.

[Eells and Skyrms, 1994] E. Eells and B. Skyrms, eds. *Probability and Conditionals*, Cambridge University Press, 1994.

[Forster and Sober, 1994] M. Forster and E. Sober. How to Tell when Simpler, More Unified, or Less Ad Hoc Theories will Provide More Accurate Predictions, *British Journal for the Philosophy of Science* 45, 1-35, 1994.

[Gaifman, 1988] H. Gaifman. A Theory of Higher Order Probabilities, in *Causation, Chance, and Credence,* Brian Skyrms and William L. Harper, eds., Dordrecht: Kluwer Academic Publishers, 1988.

[Giere, 1973] R. Giere. Objective Single-Case Probabilities and the Foundations of Statistics, in *Logic, Methodology and Philosophy of Science* IV – *Proceedings of the Fourth International Congress for Logic*, P. Suppes, L. Henkin, A. Joja, and G. Moisil (eds.), New York: North-Holland, 1973.

[Goldstein, 1983] M. Goldstein. The Prevision of a Prevision, *Journal of the American Statistical Association* 78, 817-819, 1983.

[Greaves and Wallace, 2006] H. Greaves and D. Wallace. Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility, *Mind* 115 (459), 607-632, 2006.

[Grunwald *et al.*, 2005] P. Grunwald, M. A. Pitt, and I. J. Myung, eds. *Advances in Minimum Description Length: Theory and Applications*, Cambridge, MA: MIT Press, 2005.

[Hacking, 1965] I. Hacking. *The Logic of Statistical Inference*, Cambridge: Cambridge University Press, 1965.

[Hájek, 1989] A. Hájek. Probabilities of Conditionals — Revisited, *Journal of Philosophical Logic* 18, No. 4, 423-428, 1989.

[Hájek, 1993] A. Hájek. *The Conditional Construal of Conditional Probability,* Ph.D. Dissertation, Princeton University. Available at `http://fitelson.org/conditionals/hajek_dissertation.pdf`

[Hájek, 1994] A. Hájek. Triviality on the Cheap?, in Eells and Skyrms, 113-140, 1994.

[Hájek, 2003] A. Hájek. What Conditional Probability Could Not Be, *Synthese* 137, No. 3, (December), 273-323, 2003.

[Hall, 1994] N. Hall. Back in the CCCP, in Eells and Skyrms, 141-60, 1994.

[Hausman and Woodward, 1999] D. Hausman and J. F. Woodward. Independence, Invariance and the Causal Markov Condition, *The British Journal for the Philosophy of Science* 50: 44, 521-583, 1999.

[Hempel, 1965] C. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science,* New York: Free Press, 1965.

[Hild, 1998] M. Hild. The Coherence Argument Against Conditionalization, *Synthese* 115, 229-258, 1998.

[Jaynes, 2003] E. T. Jaynes. *Probability Theory: The Logic of Science*, Cambridge: Cambridge University Press, 2003.

[Jeffrey, 1965/1983/1990] R. C. Jeffrey. *The Logic of Decision,* Chicago: Chicago University Press, 1965/1983/190.

[Jeffreys, 1961] H. Jeffreys. *Theory of Probability,* Oxford: Oxford University Press (originally published in 1939, and now in the Oxford Classic Texts in the Physical Sciences series), 1961.

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

[Keynes, 1921] J. M. Keynes. *A Treatise on Probability*, London: Macmillan, 1921.

[Kadane *et al.*, 1986] J. B. Kadane, M. J. Schervish and T. Seidenfeld. Statistical Implications of Finitely Additive Probability, in *Bayesian Inference and Decision Techniques*, Prem K. Goel and Arnold Zellner (eds.), Amsterdam: Elsevier Science Publishers, 59-76, 1986.

[Kolmogorov, 1933/1950] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitrechnung, Ergebnisse Der Mathematik*, 1933. Translated as *Foundations of Probability*. New York: Chelsea Publishing Company, 1950.

[Lange, 2000] M. Lange. Is Jeffrey Conditionalization Defective by Virtue of Being Non-Commutative? Remarks on the Sameness of Sensory Experiences, *Synthese* 123, 393-403, 2000.

[Levi, 1967] I. Levi. Probability Kinematics, *British Journal for the Philosophy of Science* 18, 197-209, 1967.

[Levi, 1974] I. Levi. On Indeterminate Probabilities, *Journal of Philosophy* 71, 391-418, 1974.

[Levi, 1980] I. Levi. *The Enterprise of Knowledge,* Cambridge: MIT Press. 1980.

[Levi, 1991] I. Levi. *Consequentialism and Sequential Choice,* in Michael Bacharach and Susan Hurley (eds.), *Foundations of Decision Theory,* Oxford: Basil Blackwell, 92-146, 1991.

[Levi, 2000] I. Levi. Imprecise and Indeterminate Probabilities, *Risk, Decision and Policy* 5, 111-122, 2000.

[Lewis, 1973] D. Lewis. *Counterfactuals,* Blackwell and Harvard University Press, 1973.

[Lewis, 1976] D. Lewis. Probabilities of Conditionals and Conditional Probabilities, *Philosophical Review* 85, 297-315, 1976.

[Lewis, 1980] D. Lewis. A Subjectivist's Guide to Objective Chance, in *Studies in Inductive Logic and Probability*, Vol II., University of California Press, 263-293, 1980; reprinted in *Philosophical Papers Volume II,* 1986, Oxford: Oxford University Press.

[Lewis, 1986] D. Lewis. Probabilities of Conditionals and Conditional Probabilities II, *Philosophical Review* 95, 581-589, 1986.

[Lewis, 1999] D. Lewis. *Papers in Metaphysics and Epistemology*, Cambridge: Cambridge University Press, 1999.

[Maher, 1992] P. Maher. Diachronic Rationality, *Philosophy of Science* 59, 120-141, 1992.

[Maher, 2007] P. Maher. Explication Defended, *Studia Logica* 86, 331-341, 2007.

[McGee, 1989] V. McGee. Conditional Probabilities and Compounds of Conditionals, *The Philosophical Review* 98, 485-541 1989.

[McGee, 1994] V. McGee. Learning the Impossible, in Eells and Skyrms, 179-99, 1994.

[Pearl, 2000] J. Pearl. *Causality: Models, Reasoning, and Inference,* Cambridge: Cambridge University Press, 2000.

[Popper, 1959a] K. Popper. The Propensity Interpretation of Probability, *British Journal of the Philosophy of Science* 10, 25–42, 1959.

[Popper, 1959b] K. Popper. *The Logic of Scientific Discovery*, Basic Books, 1959.

[Ramsey, 1931] F. P. Ramsey. General Propositions and Causality, in R. B. Braithwaite, ed., *The Foundations of Mathematics and Other Logical Essays,* Routledge; also in *Philosophical Papers,* ed. D. H. Mellor, Cambridge: Cambridge University Press, 1931.

[Reichenbach, 1956] H. Reichenbach. *The Direction of Time*, Berkeley, University of California Press, 1956.

[Renyi, 1970] A. Renyi. *Foundations of Probability*, Holden-Day, Inc. 1970.

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

[Salmon, 1980] W. Salmon. Probabilistic Causality, *Pacific Philosophical Quarterly* **61**, 50-74, 1980.

[Schwarz, 1978] G. Schwarz. Estimating the dimension of a model, *Annals of Statistics* 6 (2), 461-464, 1978.

[Seidenfeld *et al.*, 1998] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. "Non-Conglomerability for Finite-Valued, Finitely Additive Probability", *Sankhya* Series A, Vol. 60, No. 3, 476-491. 1998.

[Seidenfeld *et al.*, 2001] T. Seidenfeld, M. J. Schervish, and J. B. Kadane. Improper Regular Conditional Distributions, *The Annals of Probability* 29, No. 4, 1612-1624, 2001.

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

[Skyrms, 1980] B. Skyrms. *Causal Necessity*, Yale University Press, 1980.

[Skyrms, 1987] B. Skyrms. Dynamic Coherence and Probability Kinematics, *Philosophy of Science* 54, No. 1 (March), 1-20, 1987.

[Skyrms, 1993] B. Skyrms. A Mistake in Dynamic Coherence Arguments? *Philosophy of Science* 60, 320-328, 1993.

[Spirtes *et al.*, 2000] P. Spirtes, C. Glymour, and R. Scheines. *Causation, Prediction, and Search*, 2nd ed. New York, N.Y.: MIT Press, 2000.

[Stalnaker, 1968] R. Stalnaker. A Theory of Conditionals, *Studies in Logical Theory*, American Philosophical Quarterly Monograph Series, No. 2, Oxford: Blackwell, 1968.

[Stalnaker, 1970] R. Stalnaker. Probability and Conditionals, *Philosophy of Science* 37, 64–80, 1970.

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

[van Fraassen, 1976] B. van Fraassen. Probabilities of Conditionals, in Harper and Hooker (eds.), *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, Vol. I, Reidel, 261-301, 1976.

[van Fraassen, 1984] B. van Fraassen. Belief and the Will, *Journal of Philosophy* 81, 235-256, 1984.

[van Fraassen, 1989] B. van Fraassen. *Laws and Symmetry*, Oxford: Clarendon Press, 1989.

[van Fraassen, 1990] B. van Fraassen. Figures in a Probability Landscape, in J. M. Dunn and A. Gupta (eds.), *Truth or Consequences*, Kluwer, 1990.

[van Fraassen, 1995] B. van Fraassen. Belief and the Problem of Ulysses and the Sirens, *Philosophical Studies* 77, 7-37, 1995.

[Wallace and Dowe, 1999] C. S. Wallace and D. L. Dowe. Minimum Message Length and Kolmogorov Complexity, *The Computer Journal* 42, No. 4, (special issue on Kolmogorov complexity), 270-283, 1999.

[Walley, 1991] P. Walley. *Statistical Reasoning with Imprecise Probabilities*, London: Chapman Hall, 1991.

[Williamson, 2002] T. Williamson. *Knowledge and Its Limits*, Oxford: Oxford University Press, 2002.