Everett and Evidence

Hilary Greaves & Wayne Myrvold

August 21, 2008

Abstract

Much of the evidence for quantum mechanics is statistical in nature. Relative frequency data summarizing the results of repeated experiments is compared to probabilities calculated from the theory; close agreement between the observed relative frequencies and calculated probabilities is taken as evidence in favour of the theory. The Everett interpretation, if it is to be a candidate for serious consideration, must be capable of doing justice to this sort of reasoning. Since, on the Everett interpretation, all outcomes with nonzero amplitude are actualized on different branches, it is not obvious that sense can be made of ascribing probabilities to outcomes of experiments, and this poses a *prima facie* problem for statistical inference. It is incumbent on the Everettian either to make sense of ascribing probabilities to outcomes of experiments in the Everett interpretation, or to find a substitute on which the usual statistical analysis of experimental results continues to count as evidence for quantum mechanics, and, since it is the very evidence for quantum mechanics that is at stake, this must be done in a way that does not presuppose the correctness of Everettian quantum mechanics. This requires an account of theory confirmation that applies to branching-universe theories but does not presuppose the correctness of any such theory. In this paper, we supply and defend such an account. The account has the consequence that statistical evidence can confirm a branching-universe theory such as Everettian quantum mechanics in the same way in which it can confirm a non-branching probabilistic theory.

In the midst of this perplexity, I received from Oxford the manuscript you have examined. I lingered, naturally, on the sentence: *I leave to the various futures (not to all) my garden of forking paths*. Almost instantly, I understood: 'the garden of forking paths' was the chaotic novel; the phrase 'the various futures (not to all)' suggested to me the forking in time, not in space.

Jorge Luis Borges, "The Garden of Forking Paths."

1 Introduction

Quantum mechanics, standardly interpreted, yields, via the Born rule, statements about the probabilities of outcomes of experiments. These probabilities are, at least in many interesting cases, different from what would be expected on the basis of classical mechanics. Moreover, we can subject the claims made by standard quantum mechanics about the probabilities of outcomes of experiments to empirical test, and the results of such tests favour quantum mechanics over classical. This sort of empirical testing of probabilistic claims forms a substantial part of the evidence we have for accepting quantum mechanics as a theory that is empirically superior to classical mechanics.

Consider, for example, Bell-inequality experiments. Here we compare the probabilistic correlations yielded by a quantum-mechanical calculation to those that could be yielded by some local hidden-variables theory. Relative frequencies of outcomes in repeated trials are compared with probabilities calculated from quantum mechanics, and with probabilities that could be yielded by a local hidden-variables theory. The fact that the observed relative frequencies closely match the quantum probabilities, and exhibit statistically significant violations of Bell Inequalities, is correctly taken to favour quantum mechanics over local hidden-variable theories. Although it is possible to lose sight of the fact in discussing the bearing of such experiments on theory, the reasoning is essentially probabilistic. *Any* sequence of outcomes of such an experiment is compatible both with quantum mechanics and with local hidden-variables theories. In particular, even if some local hidden-variables theory is correct, a sequence of outcomes is *possible* (though highly improbable) in which the relative frequencies violate the Bell Inequalities. We take the observed results to rule out the latter because the results actually obtained are astronomically less probable on the assumption of a local hidden-variables theory than they are on the assumption of quantum mechanics. Similar considerations apply to the double-slit experiment. The quantum-mechanical calculation yields a probability distribution for absorption of particles by the screen. From this can be calculated a probability for any possible pattern of absorption events. The probability will be high that the observed pattern of detection events shows bands of intensity corresponding to a diffraction pattern, but we should not lose sight of the fact that *any* pattern is *consistent* with quantum mechanics, including one that matches classical expectations. The occurrence of a pattern that is much more probable on the assumption that quantum mechanics is correct than on the assumption of classical mechanics is taken to provide empirical evidence that quantum mechanics is getting the probabilities right, or approximately so.

Any interpretation of quantum mechanics that is worthy of serious consideration is going to have to make sense of this sort of reasoning. If it can't, it runs the risk of undermining the very reasons we have for accepting quantum mechanics in the first place.

On the Everett interpretation, the quantum state vector after a typical measurement interaction is a superposition of terms on which the measurement apparatus records different outcomes. Moreover, the quantum state is taken as a complete description of physical reality, so that there is nothing that distinguishes one of these branches as uniquely real. As has often been pointed out (see, *e.g.* Albert and Loewer (1988)), this poses a problem for interpreting probabilistic statements in an Everettian context. There is no obvious sense in which one can ask what the probability is that a certain result will be *the* result of the experiment, since all possible results occur in the post-experiment state, on different branches of the superposition.

There is a danger, in discussing the Everett Interpretation, of talking as if the goal is to provide a coherent interpretation that is consistent with our experience. But if that were the goal, the Everettian would have no need of probabilities; it would suffice merely to note that, for every outcome normally regarded as possible, the theory *entailed* that that outcome would occur on some branch. The goal is actually much higher: it is incumbent upon the Everettian to provide an interpretation in which the statistical analysis of the outcomes of repeated experiments provides empirical support for the theory. This is why the apparent lack of room for probability statements in the Everett interpretation threatens to create a problem for that interpretation. The problem is not one of *deriving* the correct probabilities within the theory; it is one of either making sense of ascribing probabilities to outcomes of experiments in the Everett interpretation, or of finding a substitute on which the usual statistical analysis of experimental results continues to count as evidence for quantum mechanics.

Call this the *Everettian evidential problem.* In our opinion the best hope for meeting this challenge lies in a decision-theoretic approach. The use of decision-theoretic ideas in connection with Everettian quantum mechanics was pioneered by Deutsch (1999), and elaborated, in different ways, by Wallace (2003, 2007), Saunders (2005), and Greaves (2004, 2007a); see Greaves (2007b) for a recent survey of the approach. Deutsch's argument and the variants on it presuppose an agent who accepts Everettian quantum mechanics. In order to meet the evidential problem, we need a framework for appraising theories, including branching-universe theories, that does not presuppose the acceptance of Everettian quantum mechanics or any other theory.[1] This, after some preliminary discussion in section 2, will be laid out in section 3, and applied to branching-universe theories, such as Everett's, in section 4. Section 5 discusses and replies to objections. Our conclusion (section 6) is that the framework presented here suffices to solve the evidential problem.

2 Testing probabilistic theories

We will not be in any position to address the question of whether or not statistical data can be evidence for Everettian quantum mechanics unless we are crystal clear about how *exactly* such data can be evidence for uncontroversially probabilistic theories. We therefore start by stepping back from quantum mechanics and the Everett interpretation, and reviewing some general considerations about probability statements in physics and their evaluation in the light of experimental data.

Consider the questions:

1. A pair of fair dice is about to be tossed 24 times. Which is preferable: an offer of $1,000 if a pair of sixes comes up at least once, or an offer of $1,000 if a pair of sixes never comes up?

2. *This* pair of dice is about to be tossed 24 times on *this* table, using *this* cup, by me. Which is preferable: an offer of $1,000 if a pair of sixes comes up at least once, or an offer of $1,000 if a pair of sixes never comes up?

The first question is a purely mathematical one, or close to it. Provided that you prefer receiving $1,000 to not receiving anything, then the question is one that can be answered by calculation, and is in fact the question that was posed by the Chevalier de Méré and answered by Pascal.[2]

[1] This point has also been made by Wallace (2006).

[2] See Ore (1960) for a lucid account of this incident.

The second question is not purely mathematical; it is, at least in part, a question about the physical world. To answer it, we need to know whether there is something about the physical setup—the dice, or the way they are tossed, or the make-up of the table on which they land—that biases the results towards a pair of sixes. Questions of this type can be answered in two sorts of ways:

1. By direct empirical test. Typically this involves repeated throws of the dice, and statistical analysis of the results.

2. Theoretically. This involves a theoretical model of the setup, plus a physical theory that says something about which factors are, and which factors are not, relevant to the outcome of the tosses.

A splendid example of the latter is found in Diaconis *et al.*'s model of a coin toss (Diaconis et al., 2007). They construct a simple coin-tossing machine (crucially, the coin is not permitted to bounce upon landing), model the dynamics of the tossed coin, come to conclusions regarding the probability of landing with heads or tails up, given an initial orientation, and conclude that the coin-toss is biased towards landing with the same side facing up that it started with. These conclusions are corroborated by data from repeated trials with the machine.

Even when a theoretical model is available, empirical testing is not superfluous, as we will want to satisfy ourselves of the appropriateness of the model to the case at hand. We will not, therefore, be able to do without the first way of answering the question. It is possible to overlook this point, because such calculations are usually made on the basis of symmetry considerations, and these can create the illusion that the results are truths known *a priori*. But judgments of symmetry are judgments that certain factors are irrelevant to the outcome, and this is a matter of physics. An account of probability based *exclusively* on a Principle of Indifference will not do.

Nor can probability concepts be replaced by relative frequencies, in either actual or hypothetical sequences of experiments, though relative frequency data will often be our most important sources of information about chances, or physical probabilities. Consider, for example, a case in which balls are drawn, with replacement, from an urn containing N balls in total, of which M are black, in such a way that each ball has an equal chance of being drawn. The chance, on each draw, that the drawn ball is black, is, in this case, equal to M/N, which is also the proportion of black balls in the urn. Suppose that we perform n drawings, with replacement, and let m be the number of times in these n trials that a black ball is drawn. Then, for large n, the chance is high that the sample relative frequency m/n will be close to the proportion of black balls in the population, M/N. Moreover, if the sequence of drawings be extended without end, then, with chance 1, the sample relative frequency will converge to the single-case chance M/N. Therefore, if we are unable to examine the contents of the urn, information about its contents can be gained by successive drawings. Similar ideas are behind statistical sampling techniques; one wishes to gain information about a population by a sampling of the population, and one attempts to construct one's sampling procedure such that the chance of any individual being chosen for the sample is independent of whether or not that person has the property whose proportion in the population is to be estimated. This intimate relation between chances and relative frequencies has suggested to some that chances can be *defined* in terms of relative frequencies. In spite of their intimate relation between chance and relative frequency, the former is not eliminable in favour of the latter. Notice that in the urn model, it is necessary to stipulate that each ball has an equal chance of being drawn; it is only this stipulation that makes the proportion of black balls in the urn, M/N, equal to the chance of drawing a black ball. Nor can chances be eliminated in terms of limiting relative frequencies in infinite sequences. That the relative frequency converges to the single-case chance is not the only *logically* possible outcome of the sequences of trials; it is rather the only outcome that has nonzero *chance* (and note that one cannot identify 'zero chance' with 'impossible', since even an outcome according to which relative frequency *does* match chance is an (infinite) disjunction of zero-chance outcomes). Thus, the conclusion that the limiting relative frequency will exist, and be equal to the single-case chance, requires the use of a notion of chance distinct from the notion of limiting relative frequency.

How, then, *does* the process of confirming or disconfirming statements of probability in physics work? On our view, the best way to make sense of such confirmation involves a role for two sorts of quantities that have sometimes been called "probability." The first is degree of belief, or credence, which is subjective in the sense of being attached to an (idealized) epistemic agent. Accepting this does not entail eliminating any notion of physical probability. Among the things our epistemic agent can have degrees of belief about are the chances of experimental outcomes, which are characteristic of the experimental setup, and hence the sort of things that a physical theory can have something to say about. We test such claims by performing repeated experiments—a sequence of experiments that we regard as equivalent, or near enough, with respect to the chances of outcomes—and comparing the calculated chances with the observed relative frequencies. Conditionalization on these observations raises degrees of belief in theories whose calculated chances are near the observed relative frequency and lowers degree of belief in theories whose calculated chances are far from the observed relative frequencies. That, in short, is the story of statistical confirmation of theories with experiments construed in the usual way. Its core can be summed up by the following confirmation-theoretic principle:

CC (confirmation-theoretic role of chances). If S observes something to which theory T assigned a chance higher (lower) than the average chance assigned to that same event by rival theories, then theory T is confirmed (resp. disconfirmed) for S, relative to those theories.

Note that all three concepts—credence, or degree of belief, physical chance, and relative frequency—have important roles to play in this story. The story will be elaborated upon in section 3, below, in which we provide a set of conditions, based on Savage's axioms for decision theory, and on de Finetti's concept of *exchangeability*, that are sufficient to ensure that the agent will act as if she thinks of an experiment as having chances associated with its possible outcomes, and repeated experiments as informative about the values of those chances. This permits her to experimentally test the claims a physical theory makes about chances of outcomes.

We wish to argue that a precisely analogous story can be told if the agent thinks of experiments, not in the usual way, but as involving a branching of the world, with all possible outcomes occurring on some branch or another. We claim that the conditions we introduce remain reasonable under this supposition, and that the agent will act as if she regards branches as associated with quantities, which we will call *weights*, that play in this context a role analogous to that played by chances on the usual way of viewing things. The short version of *this* story is summed up by the principle

CW (confirmation-theoretic role of branch weights). If S observes something to which theory T assigned a branch weight higher (lower) than the average chance-or-branch-weight assigned to that same event by rival theories, then theory T is confirmed (resp. disconfirmed) for S, relative to those theories.

In particular, according to our account, the agent will regard relative frequency data from repeated experiments as informative about values of branch weights in exactly the same way that, on the usual view, they are informative about chances of outcomes. If, therefore, Everettian quantum mechanics is taken as a physical theory that makes claims about branch weights, these claims can be tested by experiment.

In the general case, the agent will have non-neglible credence in some theories in which experiments are construed, in the usual way, as chance setups, and in some in which they are construed as branch setups. The account permits both to be handled simultaneously; what are estimated via repeated experiments are quantities that are to be interpreted as being *either* physical chances or physical branch-weights.

The framework will take as its starting point the notion of preferences between wagers on outcomes of experiments. This may seem an odd place to start.[3] We will first lay out a set of conditions on preferences between wagers, based on Savage's axioms, which suffice for a representation theorem, Theorem 1, according to which an agent's preferences can be represented as maximizing expected utility. On the usual interpretation, this expected utility is a weighted average of utilities across alternative epistemic possibilities, with the weighting function representing the agent's degrees of belief in these alternative possibilities. We will argue that the constraints on the agent's preferences are reasonable, also, if the agent thinks of experiments as branching events; in this case the weighting function becomes what Greaves (2007a) has called a 'quasi-credence' function. We then argue that, upon learning the results of experiments, the agent ought to update this credence-or-quasi-credence function in a manner equivalent to Bayesian conditionalization (Theorem 2). We can then take on board the de Finetti representation theorem (Theorem 3), which shows that, for an exchangeable sequence of experiments, the agent's credence-or-quasi-credence function is a weighted average of certain extremal functions that, as we will argue, can, under certain circumstances, be thought of as objective chances-or-branch-weights associated with outcomes of experiments. The weighting function (called μ in Theorem 3), under these circumstances, represents the agent's degrees of belief about which set of chances-or-branch-weights is correct. This opens the way for repeated experiments to be informative about the values of these chances-or-branch-weights: updating on observed outcomes of experiments updates the μ-function.

Some remarks on the relationship of the present paper to the existing Everettian literature are in order; these occupy the remainder of this section.

The account of decision-making and empirical confirmation of branching

[3]It is perhaps worth noting that this is where modern probability theory started, too. What we now call the mathematical theory of probability has its origins in the Fermat-Pascal correspondence (reprinted in Smith (1959)), and in the treatise of Huygens (1660). Modern readers may be surprised that these authors never calculate what we would call a probability. They are concerned, instead, with the values of wagers (expectation values, in modern parlance). It was Jacob Bernoulli's *Ars Conjectandi* that, 50 years later, introduced probabilities into the theory. "Before Bernoulli, the mathematics of games of chance had been developed by Pascal, Fermat, Huygens, and others largely without using the word (or concept of) 'probability'" (E.D. Sylla, "Preface" to Bernoulli (2006)).

theories that is defended in this paper is the same as that proposed in Greaves (2007a). The main difference between the two papers is that the present paper offers arguments (in the form of representation theorems) for two key claims that were taken as basic assumptions in Greaves (2007a). Firstly, in Greaves (2007a) it was *assumed* that, in the branching case, decisions are to be made via maximizing a weighted mean of utilities of rewards-on-branches. The present paper, in contrast, lays out a set of (Savage-style) axioms constraining rational preferences between wagers in a branching context, and spells out how the claim concerning maximization of expected utility (MEU) follows via Savage's representation theorem. While these theorems themselves are not new, their applicability to the branching case has not previously been discussed in any detail. Secondly, in Greaves (2007a) it was also *assumed* that the agent's quasi-credence function satisfies the two principles PC and PW: Principal Principle for chances, and for branch weights, respectively. The present paper points out that these claims also follow from Savage's axioms (via the De Finetti representation theorem). In both cases, our aim, in highlighting the applicability of these representation theorems to the branching case, is to shift the locus of discussion from the MEU and PW claims themselves to the axioms: if the account of rational decision-making and/or confirmation advocated here and in Greaves (2007a) is not correct then it must be that one or more of the axioms is not correct, and we urge objectors to identify which axiom they think this is.

We remark also on the relationship of the present paper to the representation theorems proved by Deutsch (1999) and Wallace (2003, 2007). Two points are worthy of note. (i) The Deutsch-Wallace approach aims to derive the Born Rule from the 'non-probabilistic' part of Everettian quantum mechanics: that is, it seeks to prove that, conditional on the truth of Everettian quantum mechanics and the given initial state for a given measurement, the rational agent's betting quotients *must* equal the corresponding amplitude-mod-squares. This is not something we claim to do in the present paper. In this respect, Deutsch and Wallace's claims are stronger than (but consistent with) ours. (ii) The decision theories developed by Deutsch and Wallace assume the truth of quantum mechanics (specifically, Everettian quantum mechanics). This means that they are not general enough to address the evidential problem. The axioms we adopt in the present paper, in contrast, are much more theory-general.

3 The Framework

In this section, we present a simplified framework, not meant to be a model for all decisions, but rather, applicable to a limited class of decisions, involving payoffs contingent on outcomes of experiments. We apply Savage's axioms to this restricted setting (Savage, 1972). These are intended to be thought of as rationality constraints on an agent's preferences between wagers.

Suppose, therefore, that we have a set of possible experiments. Associated with an experiment A there is a set S^A of possible outcomes. We do not assume that the outcome space is finite or even countable. We assume a set of payoffs, which are in the first instance the objects of our agent's preferences, and that there is a set $\mathcal{F}^A$ of subsets of S^A (the *wagerable* subsets of S^A) with which we can associate payoffs. $\mathcal{F}^A$ will be assumed to be closed under intersections, complements, and unions. An association of payoffs with the elements of a finite partition $\{F_i|\, i = 1, ..., n\}$ composed of elements of $\mathcal{F}^A$ will be called a *wager*. It will sometimes be helpful to imagine these payoffs as sums of money paid by a bookie to an agent who accepts the wager. But the framework is not limited to such cases. In particular, we allow for preferences between states of affairs that do not differ with respect to any effect on the agent (*e.g.* a sum of money paid to someone else), including states of affairs in which the agent is not present. It is not irrational to accept a wager that pays a large sum of money to your heirs in the event of your death!

To make things simpler, we will assume that, for any experiment A and any finite partition Π^A of S^A, any assignment of payoffs to elements of Π^A is a possible wager. This means that the agent is indifferent about the outcomes of experiments for their own sake, and has preferences only in so far as these outcomes lead to further consequences. One can, of course, imagine situations in which this condition does not obtain, but what matters, for our purposes, is that there is a sufficiently rich set of experiments and outcomes that are such that this condition is, for all intents and purposes, realized.

A wager $\mathbf{f}$ on a partition Π^A of the outcome space of an experiment A can be represent by the function that associates payoffs with the outcomes of A. If A is an experiment, $\{F_i|\, i = 1, ..., n\}$ a partition of S^A, and $\{a_i\}$ a set of payoffs, we will write $[F_i \to a_i]$ for the wager on A that pays a_i on outcomes in F_i. We will also write $[F \to a, \neg F \to b]$ for $[F \to a, (S^A - F) \to b]$.

Performing one experiment may preclude performance of another. We assume that there is a relation of compatibility on the set of experiments. Bets on compatible experiments can be combined. For any two compatible experiments A, B, there is a third experiment C, with outcome space $S^A \times S^B$, such that outcome $(s, t) \in S^C$ occurs iff $s \in S^A$ and $t \in S^B$ occur. For any subset $F \subseteq S^A$, there will be a corresponding subset $F \times S^B$ consisting of all $(s,t) \in S^A \times S^B$ such that $s \in F$. For notational convenience, we will occasionally ignore the distinction between F and $F \times S^B$, and will write $F \cap G$ for the set of $(s,t) \in S^A \times S^B$ such that $s \in F$ and $t \in G$: that is, $(F \times S^B) \cap (S^A \times G)$.

We assume that our agent has a preference ordering $\preceq$ on the set of wagers. The following axioms, based on those of Savage (1972), are to be taken as rationality constraints on this preference ordering.

P1. a) $\preceq$ is transitive. That is, for all wagers $\mathbf{f}, \mathbf{g}, \mathbf{h}$, if $\mathbf{f} \preceq \mathbf{g}$ and $\mathbf{g} \preceq \mathbf{h}$, then $\mathbf{f} \preceq \mathbf{h}$.

b) $\preceq$ is a total ordering. That is, for all wagers $\mathbf{f}$, $\mathbf{g}$, $\mathbf{f} \preceq \mathbf{g}$ or $\mathbf{g} \preceq \mathbf{f}$.

(Note that reflexivity of $\preceq$ follows from (b)).

We define an equivalence relation $\approx$ by,

$$\mathbf{f} \approx \mathbf{g} \text{ iff } \mathbf{f} \preceq \mathbf{g} \text{ and } \mathbf{g} \preceq \mathbf{f}.$$

We define strict preference $\prec$ by,

$$\mathbf{f} \prec \mathbf{g} \text{ iff } \mathbf{f} \preceq \mathbf{g} \text{ and } \mathbf{g} \not\preceq \mathbf{f}.$$

We introduce the concept of a *null* outcome set as one that is disregarded in all considerations of desirability of wagers. Obviously, the empty set is a null set; we leave open the possibility that there might be others, regarded as by the agent as negligible in all deliberations regarding preferences between wagers. (Heuristically: in the probabilistic case, null outcomes are those to which the agent ascribes zero probability.)

Definition. Let A be an experiment, $F \in \mathcal{F}^A$. F is *null* iff, for all wagers $\mathbf{f}, \mathbf{g}$ that differ only on F, $\mathbf{f} \approx \mathbf{g}$.

The next axiom says that preferences between wagers depends only on their payoffs on the class of outcomes on which the wagers disagree. If I have wagers $\mathbf{f}, \mathbf{g}$ on an experiment A, that differ only on an outcome set F and agree (yield the same payoffs) on $S^A - F$, then I can replace them by wagers $\mathbf{f}', \mathbf{g}'$ that agree with $\mathbf{f}, \mathbf{g}$, respectively, on F, and agree with each other on $S^A - F$, without changing the preference ordering.

P2. Let A be an experiment, $F \in \mathcal{F}^A$ a set of outcomes of A, and let $\mathbf{f}, \mathbf{f}', \mathbf{g}, \mathbf{g}'$ be wagers on A such that, on F, $\mathbf{f}$ agrees with $\mathbf{f}'$ and $\mathbf{g}$ agrees with $\mathbf{g}'$, and on $S^A - F$, $\mathbf{f}$ agrees with $\mathbf{g}$ and $\mathbf{f}'$ agrees with $\mathbf{g}'$. If $\mathbf{f} \preceq \mathbf{g}$, then $\mathbf{f}' \preceq \mathbf{g}'$.

The next axiom is context-independence of preferences between payoffs. A preference for receiving b to a as a result of one wager carries over to other wagers.

P3. Let A, B be experiments, let $\mathbf{f}, \mathbf{f}'$ be wagers on A that pay a, b, respectively, on $F \in \mathcal{F}^A$, and coincide otherwise, and let $\mathbf{g}, \mathbf{g}'$ be wagers on B that pay a, b, respectively, on $G \in \mathcal{F}^B$, and coincide otherwise. If $\mathbf{f} \prec \mathbf{f}'$, then $\mathbf{g} \preceq \mathbf{g}'$, and $\mathbf{g} \approx \mathbf{g}'$ only if G is null.

For any payoff a and any experiment A, there will be a trivial wager $I^A(a)$ that pays a no matter what happens. P3 ensures that preferences between such trivial wagers are independent of the experiment performed. With this axiom in place, the preference order on wagers induces a preference order on payoffs: $a \preceq b$ iff $I^A(a) \preceq I^A(b)$ for some experiment A (hence, by P3, for all experiments).

Suppose I am given a choice between wagers:

$\mathbf{f}$: Receive \$1,000 on F, nothing otherwise.
$\mathbf{g}$: Receive \$1,000 on G, nothing otherwise.

Suppose I prefer $\mathbf{g}$ to $\mathbf{f}$. Then it is reasonable to expect that this preference would not change if some other payoff that I prefer to receiving nothing were substituted for the \$1,000. The, (assuming I like chocolate cupcakes), I should therefore also prefer prefer $\mathbf{g}'$ to $\mathbf{f}'$, where these are defined by

$\mathbf{f}'$: Receive a chocolate cupcake on F, nothing otherwise.
$\mathbf{g}'$: Receive a chocolate cupcake on G, nothing otherwise.

The next axiom is meant to capture this intuition.

P4. Let A, B be experiments, $F \in \mathcal{F}^A$, $G \in \mathcal{F}^B$. If a, b, a', b' are payoffs such that $b \prec a$ and $b' \prec a'$, and $[F \rightarrow a, \neg F \rightarrow b] \preceq [G \rightarrow a, \neg G \rightarrow b]$, then $[F \rightarrow a', \neg F \rightarrow b'] \preceq [G \rightarrow a', \neg G \rightarrow b']$.

With this axiom in place, we can define an ordering $\preccurlyeq$ on wagerable outcome sets.

Definition. For $F \in \mathcal{F}^A$, $G \in \mathcal{F}^B$, $F \preccurlyeq G$ iff there exist payoffs a, b such that $a \prec b$ and $[F \rightarrow a, \neg F \rightarrow b] \preceq [G \rightarrow a, \neg G \rightarrow b]$.

It is easy to check that $\preccurlyeq$, so defined, is a reflexive, transitive, total ordering. We define an equivalence relation $F \sim G$ as $F \preccurlyeq G$ and $G \preccurlyeq F$, and a strict order $F \prec G$ as: $F \preccurlyeq G$ and not $F \sim G$. (Heuristically: in the probabilistic case, if $F \sim G$ then F and G are regarded as equally likely by the agent.) If $F \prec G$, then G counts for more in our agent's deliberations than F. Differences in payoffs attached to G have more effect on desirability of the overall wager than differences in payoffs attached to F. If $F \sim G$, then F and G hold the same weight in our agent's deliberations.

So far, everything that has been said is compatible with the preference ordering being a trivial one: $\mathbf{f} \approx \mathbf{g}$ for all wagers $\mathbf{f}, \mathbf{g}$. This is the preference ordering of an agent who has achieved a state of sublime detachment. To exclude such a state of nirvana, we add a non-triviality axiom.

P5. There exist payoffs a, b such that b is strictly preferred to a, that is, $a \prec b$.

We want to be able to turn the qualitative relation $\preccurlyeq$ into a quantitative one. That is, we want to associate with each outcome set F a number $\alpha(F)$ such that $F \preccurlyeq G$ iff $\alpha(F) \leq \alpha(G)$. We can do this if, for every n, there is an experiment A and an n-element partition $\{F_i\}$ of S^A such that $F_i \sim F_j$ for all i, j. Assigning $\alpha(\emptyset) = 0$ and $\alpha(S^A) = 1$ then gives us $\alpha(F) = m/n$ for any union of m distinct elements of this partition. Armed with sets of outcomes on which α takes on all rational values, the fact that $\preccurlyeq$ is a total ordering gives us for any outcome-set G a real number value $\alpha(G)$.

It turns out that we can assume something a bit weaker. If we can always find experiments such that all outcomes are arbitrarily low in the $\preccurlyeq$-ordering, then we can construct n-partitions that are arbitrarily close to being equivalent, and so get a real-valued ordering function in that way. This is Savage's procedure. Thus we add one last axiom,

P6. Let $\mathbf{f}$, $\mathbf{g}$ be wagers on experiments A, B, respectively, such that $\mathbf{f} \prec \mathbf{g}$. Then, for any payoff a, there is an experiment C, compatible with both A and B, and a partition Π^C of S^C, such that, for each element $F \in \Pi^C$, if we consider the modified wager $\mathbf{f}'$ on the combination of A and C that pays a on F, and coincides with $\mathbf{f}$ otherwise, we have $\mathbf{f}' \prec \mathbf{g}$. Similarly, if we form $\mathbf{g}'$ by paying a on F and retaining $\mathbf{g}$'s payoff otherwise, then we have $\mathbf{f} \prec \mathbf{g}'$.

We now have all the conditions we need for a representation theorem.

Theorem 1 *(Savage). If the preference ordering $\preceq$ satisfies P1 – P6, then there exists a utility function u on the set of payoffs (unique up to positive linear transformations), a function α (unique up to a scale factor), which takes as arguments wagerable subsets of experimental outcome-spaces, and a* *function U on the set of possible wagers, such that, for any experiment A, wagerable partition $\{F_i |\, i = 1, ...n\}$ of S^A, and wager $\mathbf{f} = [F_i \rightarrow a_i]$,*

$$U(\mathbf{f}) = \sum_{i=1}^{n} \alpha(F_i)\, u(a_i)$$

and, for all wagers $\mathbf{f}$, $\mathbf{g}$, $U(\mathbf{f}) \leq U(\mathbf{g})$ iff $\mathbf{f} \preceq \mathbf{g}$.

Theorem 1 says that our agent's judgments, if they satisfy P1–P6, are as if the agent is maximizing expected utility with u giving the utilities attached to payoffs, and the α function acting as if it represents degrees of belief in the outcomes of experiments. See Savage (1972) for proof.

3.1 Learning

Our agent may revise her judgments about wagers on future experiments upon learning the results of past experiments: she may learn from experience. Suppose an experiment A is to be performed, and that our agent is to learn which member of an n-element partition $\{D_i^A\}$ the outcome of A falls into, after which she will be given a choice between wagers $\mathbf{f}$ and $\mathbf{g}$, defined on a partition $\{E_j^B \mid j = 1, ..., m\}$ of S^B. Her choice of wager on B may, in general, depend on the outcome of A. There are 2^n strategies that she can adopt, specifying, for each D_i^A, whether her choice would be $\mathbf{f}$ or $\mathbf{g}$ were she to learn that outcome of A was in D_i^A. Her choice of which strategy to adopt is equivalent to a choice among a set of 2^n wagers on the combined outcome of A and B. Each such wager consists of specifying, for each i, whether the payoff on $D_i^A \cap E_j^B$ will be $\mathbf{f}$'s payoff on E_j^B for every j, or $\mathbf{g}$'s payoff on E_j^B. Our agent's preference ordering on wagers therefore induces a preference ordering on updating strategies.

We wish to consider changes of preference that can be regarded as pure learning experiences. This means: changes that do not involve a re-evaluation of the agent's prior judgments, and come about solely as result of acquiring a new piece of information. We do not claim that no other change of preference is rational; the agent may re-assess her judgments and revise them as a result of mere cogitation. For changes that are *not* of this sort, the following axiom is a reasonable constraint (and may even be taken as part of what one *means* by a 'pure learning experience').

P7. During pure learning experiences, the agent adopts the strategy of updating preferences between wagers that, on her current preferences, she ranks highest.

This preferred updating strategy is easy to characterize.

Theorem 2 *Define the updated utility that assigns the value*

$$U_i^A(\mathbf{f}) = \sum_{j=1}^{m} \alpha_i^A(E_j^B)\, u(f_j),$$

to the wager $[E_j^B \rightarrow f_j]$ *on* B*, where* α_i^A *is defined by*

$$\alpha_i^A(E_j^B) = \frac{\alpha(D_i^A \,\cap\, E_j^B)}{\alpha(D_i^A)}.$$

for non-null D_i^A*. The strategy that recommends, upon learning that the outcome of experiment* A *is in* D_i^A*, that subsequent choices of wagers be made on the basis of* U_i^A*, is strictly preferred to any other updating strategy.*

Theorem 2 says that the strategy that ranks highest in our agent's preference ordering is the strategy equivalent to updating by conditionalization. See Appendix for proof.

3.2 Repeatable experiments

We are interested in repeatable experiments. Now, no two experiments are exactly alike (for one thing, they occur at different places or different times, which in practice means that the physical environment is different in *some* respect). But our agent might regard two experiments as essentially the same, at least with respect to preferences between wagers on outcomes. Suppose we have a sequence $\mathcal{A}$ of mutually compatible experiments $\{A_1, A_2, ...\}$, with isomorphic outcome spaces. For ease of locution, we will simply identify the outcome spaces, and speak as if two elements of the sequence can yield the same outcome. If, for every composite wager formed from independent wagers on each of a finite subsequence $\mathcal{E}$ of $\mathcal{A}$, the value of the wager is unchanged if the payoffs attached to any two elements of $\mathcal{E}$ are switched, we will say that the sequence $\mathcal{A}$ is a sequence of *repeatable* experiments. (De Finetti called such a sequence an *exchangeable* sequence.) Note that repeatability/exchangeability is, properly speaking, a characteristic, not of the sequence of experiments, but of an agent's preference ordering over wagers on the outcomes of the experiments, and reflects judgments that the agent makes about which factors are irrelevant to the value of a wager. Note that our agent's judgments about wagers on experiments in a sequence of repeatable experiments need not be independent of each other. Knowing the outcome of one experiment might be relevant to judgments about the value of wagers on other members of the sequence.

We can use de Finetti's representation theorem to characterize the α-functions, and, hence, the utility functions U, on which a sequence will be exchangeable. Among utility functions that make $\mathcal{A}$ an exchangeable sequence there are some that make wagers independent of each other, in the sense that knowing the outcome of some subset of experiments in the sequence makes no difference to the evaluation of wagers on the other elements of the sequence. The de Finetti theorem specifies the form of these utility functions, and says that any utility function that makes $\mathcal{A}$ exchangeable is a mixture of such utilities.

First, some definitions that will facilitate stating the theorem. Let $\mathcal{A}$ be a sequence of mutually compatible experiments, and let $\Pi^A = \{F_i \,|\, i = 1, ..., n\}$ be a partition of their common outcome space S^A. For any finite subsequence $\mathcal{E}$ of $\mathcal{A}$, let $\mathfrak{A}_{\mathcal{E}}$ be the composite experiment consisting of elements of $\mathcal{E}$. If $\mathcal{E}$ is an m-element subset, the outcome space of $\mathfrak{A}_{\mathcal{E}}$ is $S^A \times ... \times S^A$ (m times). Form the partition Σ of this outcome space whose elements (n^m of them) correspond to specifying, for each experiment $A_i \in \mathcal{E}$, which member of the partition Π^A the outcome of A_i falls into. For each $s \in \Sigma$, let $\mathbf{k}(s)$ be the vector $(k_1, k_2, ..., k_n)$, where k_i specifies how many times an outcome in F_i occurs in s. For example, if $m = 10$, and Π^A is a two-element partition $\{F_1, F_2\}$, one element of Σ would be

$$s = (1, 2, 2, 2, 1, 1, 2, 2, 2, 1),$$

and we would have $k_1(s) = 4, k_2(s) = 6$. Note that we must have

$$\sum_{i=1}^{n} k_i = m.$$

We will be interested in wagers on which payoffs are paid independently on elements of $\mathcal{E}$; that is, wagers $\mathbf{f}$ composed of wagers $\mathbf{f}_j = [F_i \rightarrow a_{ji}]$ on $A_j \in \mathcal{E}$. The sequence $\mathcal{A}$ is exchangeable if, for every finite subsequence $\mathcal{E}$, and any such composite wager $\mathbf{f}$ on $\mathfrak{A}_{\mathcal{E}}$, the value of $\mathbf{f}$ is unchanged by permutations of the component wagers $\mathbf{f}_j$.

Let Λ_n be the $(n-1)$-dimensional simplex consisting of vectors $\lambda = (\lambda_1, ..., \lambda_n)$ satisfying the constraint:

$$\sum_{i=1}^{n} \lambda_i = 1.$$

For any $\lambda \in \Lambda_n$, we can define an α-function,

$$\alpha_\lambda(s) = \lambda_1^{k_1(s)}\, \lambda_2^{k_2(s)} ... \lambda_n^{k_n(s)}.$$

It is easy to check that the utility functions that assign values to wagers on m-member subsets of $\mathcal{A}$ by

$$U_\lambda(\mathbf{f}) = \sum_{s \in \Sigma} \alpha_\lambda(s)\, u(f_s)$$

are ones on which $\mathcal{A}$ is exchangeable, and, moreover, are ones on which elements of the sequence are independent of each other. What de Finetti showed is that *any* utility function on which $\mathcal{A}$ is exchangeable can be written as a mixture of such functions.

Theorem 3 *(De Finetti, 1937) Let $\mathcal{A}$ be an exchangeable sequence of experiments, $\{F_i \mid i = 1, ..., n\}$ a partition of their common outcome space S^A. Then there is a measure μ on Λ_n such that, for any wager $\mathbf{f}$ on the outcomes of a finite subsequence $\mathcal{E}$ of experiments in $\mathcal{A}$,*

$$U(\mathbf{f}) = \int_{\Lambda_n} d\mu(\lambda)\, U_\lambda(\mathbf{f}).$$

We are now close to having all the conditions required for our agent to take relative frequency of results of past experiments in an exchangeable sequence as a guide to future preferences between wagers. Close, but not quite there. Consider an agent who initially bets at even odds on a coin toss. Suppose, now, that the coin is tossed one hundred times, with heads coming up each time. We would regard it as reasonable for the agent to favour heads on the next toss: she should prefer a reward on heads to the same reward on tails. It is, however, compatible with all the conditions above, including that she treat successive coin tosses as exchangeable, that our agent resist learning from past experience, and continue to bet at even odds. We therefore add a condition that her preferences be non-dogmatic.

P8. For any exchangeable sequence $\mathcal{A}$, the measure μ appearing in the de Finetti representation should not assign measure zero to any open subset of Λ_n.

We now have learning from experienced within an exchangeable sequence. As an example, consider a repeated coin flip. Since we have only two possible outcomes for each flip, Λ_n is just the unit interval $[0, 1]$, and the extremal α functions can be characterized by a single parameter λ. These extremal α-functions are those that assign, to a sequence s of N flips containing m heads and $n = N - m$ tails, the value

$$\alpha_\lambda(s) = \lambda^m (1 - \lambda)^n.$$

Suppose that the measure μ is represented by a density function $\mu(\lambda)$.

$$\alpha = \int_0^1 d\lambda\, \mu(\lambda)\, \alpha_\lambda$$

After observing a sequence s of N tosses containing m heads and n tails, our agent updates the α-function she uses to evaluate subsequent wagers by conditionalization,

$$\alpha \rightarrow \alpha_s,$$

which is equivalent to updating the density function μ via

$$\mu \rightarrow \mu_s,$$

where

$$\mu_s(\lambda) \; \propto \; \lambda^m (1-\lambda)^n \, \mu(\lambda).$$

The function

$$l(\lambda) = \lambda^m (1-\lambda)^n$$

is peaked at $\lambda = m/N$, which is the relative frequency of heads in the observed sequence s. Moreover, it is more sharply peaked (with a width that goes as $1/\sqrt{N}$), the larger the observed sequence. Thus, if our agent's initial α-function is non-dogmatic, for sufficiently large N the density μ will end up concentrated on an interval around the observed relative frequency, with width of order $1/\sqrt{N}$.

Strict exchangeability is a condition that will rarely be satisfied for agents with realistic judgments about wagers. The agent might not be *completely* certain that differences between elements of the sequence $\mathcal{A}$ ought to be regarded as irrelevant. If they are successive throws of a die, for example, our agent might not completely disregard the possibility that some observable feature of the environment is relevant to the outcomes of the die. Her α-function, accordingly, will be a mixture of one on which the sequence is exchangeable, and others containing correlations between the elements of $\mathcal{A}$ and the results of other possible experiments. There are generalizations of the de Finetti representation theorem that encompass such situations. Not surprisingly, they have the result that the agent can learn which experiments she ought to take as correlated and which she ought to take as independent, and may converge towards a judgment of exchangeability regarding a sequence of possible experiments. See Diaconis and Freedman (1980), Skyrms (1984, Ch. 3), and Skyrms (1994) for discussions of such generalizations.

3.3 On the notion of physical chance

The de Finetti representation theorem shows that an agent whose degrees of belief make a sequence of coin tosses an exchangeable sequence will bet in exactly the same way as someone who believes that there is an objective chance, perhaps imperfectly known, for each toss to come up heads, and who has degrees of belief concerning the value of this chance, which mesh with her degrees of belief concerning outcomes of the tosses in the way prescribed by Lewis' Principal Principle (Lewis, 1980). Furthermore, if our agent's degrees of belief are nondogmatic, she will, upon learning the results of an initial finite sequence of tosses, update her betting preferences in exactly the same way as someone who takes these tosses to be informative about the chance of heads on the next toss. This has been taken by some—and was so taken by de Finetti— to indicate that the notion of objective chance is eliminable. There is another way to look at it, however: the agent's degrees of belief are, implicitly, degrees of belief about objective chances. An extremal α-function α_λ represents a chance distribution on which the chance of obtaining a result in F_i is λ_i, and the mixture represents the agent's degrees of belief about which of these functions give the actual chances.[4]

Should we, then, in some circumstances at least, ascribe beliefs about objective chances to agents? Note that, even if we start with the idea that probabilities are subjective, we are not thereby committed to denying that some probability assignments are better adapted to the world than others. De Finetti famously declared that the only criterion of admissibility of probability assignments is that of *coherence*; all probability assignments "are admissible assignments: each of these evaluations corresponds to a coherent opinion, to an opinion legitimate in itself, and every individual is free to adopt that one of these opinions which he prefers, or, to put it more plainly, that which he *feels*" (De Finetti, 1980, p. 64). Such language suggests that all probability assignments are equally valuable, but note that de Finetti is careful not to say that. Once an agent has adopted a probability assignment, she will not freely exchange it for any other. Nor will an agent always regard her own judgments to be the best. Suppose that Alice and Bob both have degrees of belief on which a certain sequence of experiments is exchange-

[4]There is an analogy here with the relationship of the principle of maximizing expected utility (MEU) to the Savage representation theorem. The Savage theorem shows that, given certain constraints on preferences, there will be a credence and utility functions according to which her preferences satisfy MEU. We are, in effect, using MEU to ascribe credences and utilities to the agent. Similarly, the de Finetti theorem shows that, given certain constraints, the agent acts as if she has credences about chances, credences that satisfy the Principal Principle. One could say: it is via this Principle that we ascribe beliefs about chances to the agent.

able, and that their priors are the same, or close enough that differences are negligible. Suppose Alice learns the result of the first 100 elements of the sequence, and Bob does not, and that they are both offered wagers on the result of the 101st. Unless Bob has zero degree of belief that learning what Alice knows would affect his judgment about the wager he is about to undertake, coherence requires that strictly prefer betting according to Alice's judgments to betting according to his own current judgments, if offered the choice. He does *not* regard all assignments of probability as equally valuable, and does not even rank his own highest.

Suppose, now, that there is a sequence $\mathcal{A}$ of experiments that Bob judges to be exchangeable, and that there are no other experiments except those in $\mathcal{A}$ that he takes to be relevant to elements of the sequence, and suppose his preferences are non-dogmatic. Then, if offered the opportunity to accept or reject wagers on an element A of the sequence, he would certainly prefer to have knowledge of outcomes of other elements of the sequence that have already been performed. Furthermore, if there are elements of the sequence that have not been performed, but could have been, he would prefer that they had been performed, because knowledge of the outcomes of these would improve his betting situation. There will, however, typically be no experiments that either have or could have been performed that would lead him to certainty regarding the outcome of A. However, he is certain that there is some probability function over the potential outcomes of A to which his degrees of belief, and those of any other agent who judged the sequence exchangeable and was non-dogmatic, would converge, were they to learn the results of sufficiently many other members of the sequence.

Suppose that on Bob's credences, the results of experiments not in the sequence $\mathcal{A}$ are irrelevant to experiments in $\mathcal{A}$. Then, the extremal α-functions are invariant under conditionalization on the results of any experiment that has been or could have been performed prior to betting on a given element of $\mathcal{A}$. They are, in this sense, regarded by Bob as candidates for being the maximally well-informed, or optimal betting strategy. He does not currently know which one of them is in fact optimal, but his current betting preferences are epistemically weighted averages reflecting his current degrees of belief about what the optimal strategy is. The optimal strategy is not subjective, in the sense of being the betting strategy of any agent. It is something that Bob regards as optimal for bets on a certain class of experimental setups. Furthermore, when he conditionalizes on the results of elements of the sequence, he learns about what the optimal strategy is, and he is certain that any agent with non-dogmatic priors on which the sequence of experiments is exchangeable will converge to the same optimal strategy. If this is not the same as believing that there are objective chances, then it is something that serves the same purpose. Rather than eliminate the notion of objective chance, we have uncovered, in Bob's belief state, implicit beliefs about chances—or, at least, about something that plays the same role in his epistemic life.

To generalize beyond the case of exchangeability: suppose that Bob has degrees of belief regarding the outcome of an experiment A, which can be represented as mixtures of probability functions that he regards as states of maximal accessible knowledge, in the sense of being invariant under conditionalization on results of all experiments that either actually have or could have been performed prior to A, and suppose that we can show that, with probability one, Bob's beliefs would converge to one of these, given a sufficient body of information of the sort that could be accessible to an agent about to bet on the outcome of A. Then Bob's preferences between wagers are as if he thinks that one of these extremal, maximally informed probability distributions is the correct chance distribution, and his preferences reflect degrees of belief about what the chance distribution is.[5]

Presumably, the physics of an experimental setup is relevant to which betting strategy on outcomes of the experiment is optimal. Bob may formulate theories about what the optimal strategy is for a given experimental setup. Experiments that he regards as informative about these optimal strategies will accordingly raise or lower his degrees of belief in such theories. One sort of theory would be one in which the dynamical laws are stochastic, invoking an irreducible chance element. The theory could also have deterministic dynamics. Though such a theory will map initial conditions into outcomes of experiments, it might nevertheless be the case that the maximal *accessible* information (confined to learning the results of all experiments that have or could have been performed, prior to the experiment on which the wager is placed) falls short of information sufficient to decide with certainty between experimental outcomes. This is the case with the Bohm theory. Though the theory is deterministic, it is a consequence of the theory that no agent can have knowledge of particle positions that would permit an improvement over betting according to Born rule probabilities. In this context, these maximally informed degrees of belief play the role of objective chances.

Lewis remarked, of the notion of objective chance, "Like it or not, we have this concept" (Lewis, 1980, p. 269). To which we might add: like it or not, an agent with suitable preferences acts as if she believes that there are objective chances associated with outcomes of the experiments, about which she can learn, provided she is non-dogmatic. This, together with the assumption that physical theories may have something to say about these chances, is all

[5]This discussion is heavily indebted to that found in Chapter 3 of Skyrms (1984). See also Skyrms (1994)

we require for our account of theory confirmation. There may be more to be said about the nature and ontological status of such chances, but, whatever more is said, it should not affect the basic picture of confirmation we have sketched.

Though the notion of physical chance is not reducible either to epistemic probability or frequency, the three are intimately related. An agent who updates her epistemic probabilities by conditionalizing on the results of repeated experiments will take the relative frequencies of outcomes in these experiments as evidence about the values of physical chances. In this way theories that say something about physical chances are confirmed or disconfirmed by experiment. Note that we have not needed to pass to an infinite limit to achieve such confirmation. Nor is there any need for a substantive additional assumption such as "Assume your data are typical." It is a consequence of conditionalizing on the data that degree of belief is raised in theories that posit chances that are close to the observed relative frequencies and lowered in theories that posit chances that are far from the observed relative frequencies.

4 The Garden of Forking Paths

Suppose, now, that our agent, having read Borges' "The Garden of Forking Paths," (Borges, 1941, 1962) thinks of an experiment as an event in which the world divides into branches, with each outcome occurring on some branch. On each of the branches is a copy of herself, along with copies of everyone else in the world, and each payoff is actually paid on those branches on which the an outcome associated with that payoff occurs. How much of the foregoing analysis would have to be revised?

We claim: none of it. The Savage axioms are requirements on the preferences of a rational agent, whether the agent conceives of an experiment in the usual way, with only one outcome, or as a branching occurrence, with all of the payoffs actually paid on some branch or another. The reader is invited to go back and reconsider the axioms in this light. (We will discuss some possible objections to this claim in section 5.)

Reinterpreting experiments in this way, however, does force a reinterpretation of the α-functions that appear in the representation of the agent's preferences. The reason is that on a branching interpretation of experiments, $\alpha(F)$ cannot in general be interpreted as degree of belief that the outcome of the experiment will lie in the set F: our agent may have degree of belief 1 that each outcome associated with a non-null subset of S^A will occur (on some branch), but still in general $\alpha(F) < 1$. What we *can* say, on the basis of the way it (still) feeds into the maximization of expected utility formula, is that the function $\alpha(F)$ is a measure of the weight the agent attaches in her deliberations to branches having outcomes in F.

What the De Finetti representation shows (now) is that, for an exchangeable sequence, the agent's α-function will have the form of degrees of belief concerning optimal branch weights, where these 'branch weights' play the role of physical chances in her deliberations. When our agent updates her preferences by conditionalization on experimental results, she will take the results of previous experiments in an exchangeable sequence as informative about branch weights (rather than about chances).

Ordinary quantum mechanics consists of the Hilbert space framework, plus interpretive rules that tell us how to associate operators with experimental setups and state vectors (or density operators) with preparation procedures, plus the Born rule, which tells us to interpret the squares of amplitudes as chances of outcomes of experiments. It is this latter rule that gives the theory much of its empirical content; theories that make claims about physical chances are confirmed or disconfirmed in the manner described in the previous section.

Now consider Everettian quantum mechanics as a theory that retains the Hilbert space framework, the same associations of operators with experimental setups and state vectors or density operators with preparation procedures, but replaces the Born rule with the rule: the squares of amplitudes are to be interpreted, not as chances of outcomes, but as branch weights. The calculated values can be compared with the results of experiments, and Everettian quantum mechanics is confirmed in much the same way as quantum mechanics with Born-rule chances is.

On this view, we (as agents who are agnostic about whether or not our world is a branching one) should be taking relative frequency data as informative about quantities that are *either* physical chances *or* physical branch weights. A hypothesis that makes claims about physical branch weights is confirmed by the data to precisely the same extent as a hypothesis that attributes the same numerical values to chances. As with chance, there may be more to said about the nature and ontological status of these branch weights, but such a further account is not expected to affect the basics of how branch-weight theories are confirmed.

We close this section with two examples, intended to clarify and fix ideas by showing how this account works in two particular cases.

4.1 Example: the unbiased die

Recall the question with which we started: whether, on 24 tosses of a pair of fair dice, it is better to receive $\$1,000$ if a pair of sixes come up at least once, or to receive $\$1,000$ if a pair of sixes never come up. We now consider this question in a branching context.

There is a continuum of ways in which the dice can land, but we are interested only in which faces are up when the dice come to rest. We therefore partition the outcome space into the $36^{24} \approx 2.24 \times 10^{37}$ classes corresponding to distinct sequences of results. Suppose that these classes of outcomes are all regarded by our agent as equivalent, with respect to wagers—she will not change her estimation of the value of a wager on this experiment upon permutation of the payoffs associated with elements of the partition. Then she ought to prefer a wager $\mathbf{g}$ that pays $\$1,000$ on the branches on which a pair of sixes does not occur, and nothing on all branches on which this does occur, to a wager $\mathbf{f}$ with the payoffs reversed. Why? Because there are $35^{24} \approx 1.14 \times 10^{37}$ sequences of results on which a pair of sixes never occurs, and only $36^{24} - 35^{24} \approx 1.10 \times 10^{37}$ on which at least one pair of sixes does occur. The wager $\mathbf{f}$ can be converted via a permutation of payoffs into a wager $\mathbf{f}'$ in which the $\$1,000$ is received on $36^{24} - 35^{24}$ elements of our partition on which a pair of sixes does not occur, and nothing is paid on the remaining branches. By assumption, this does not change the value of the wager, and so $\mathbf{f} \approx \mathbf{f}'$. We can obtain $\mathbf{g}$ from $\mathbf{f}'$ by giving a $\$1,000$ reward on each of the remaining branches — corresponding to approximately 4×10^{35} elements of our partition — on which a pair of sixes does not occur. If it is better for the agents on those branches to receive $\$1,000$ than to receive nothing, then, by P2, we should regard $\mathbf{g}$ as preferable to $\mathbf{f}'$, and hence, to $\mathbf{f}$.

4.2 Example: the biased die

Suppose that our agent has available to her records of the outcomes of a great many previous rolls of the die, and examination shows that, though one of them displays the behaviour expected of a fair die, the other has shown a 6 in a fraction of outcomes significantly higher than 1/6. On the ordinary view of a die toss, we would say that it should be possible for sufficient data of this sort to reverse her estimates of the values of the wagers $\mathbf{f}$ and $\mathbf{g}$, and come to prefer $\mathbf{f}$ to $\mathbf{g}$. It does not take a huge bias to reverse this preference. If, for example, one die is unbiased, and the other has a chance of 6/35 of showing a six on any given toss, then the chance that a pair of sixes shows up at least once in 24 tosses is approximately 0.501, and $\mathbf{f}$ is the marginally better wager.

On a branching view also, if our agent has exchangeable, non-dogmatic prior preferences about wagers on the dice tosses, then a sufficiently large number of tosses showing bias will lead her to prefer **f** to **g**. She regards the statistical evidence as informative about branch weights and concludes that, though there is a greater number of sequences of possible outcomes of 24 dice tosses in which a pair of sixes never occurs, the set of branches on which a pair of sixes comes up at least once has a higher total weight. That is, it is better to reward her successors on that set of branches, than on its complement.

A simpler example will give the flavour of this reasoning. Suppose that our agent is initially sure that a coin is either biased two-to-one in favour of heads, or two-to-one in favour of tails, with her degrees of belief evenly divided between these two alternatives. That is, she believes that the coin either produces, on each toss, branches with total weight 2/3 on which it lands heads and branches with total weight 1/3 on which it lands tails, or branches with total weight 1/3 on which it lands heads and total weight 2/3 on which it lands tails. Suppose that the coin is to be tossed twice, and that after learning the result of the first toss, she will be given the choice between receiving $1,000 if the second toss lands heads, and $1,000 if the second toss lands tails. She resolves to bet on heads on the second toss if the first toss is heads, and on tails if the first toss is tails. She reasons as follows. If the coin is biased towards heads, then it is better to make the second bet on heads. On her strategy, this will happen on weight 2/3 of branches, with the wrong bet being made on weight 1/3. Similarly, if the coin is biased towards tails. Her estimation of a strategy is an epistemically weighted mean of its value if the coin is biased towards heads, and its value if the coin is biased towards tails. The strategy she has resolved to follow is the one with the highest expected value.

If she is to be coherent, and if she is to follow this strategy upon learning the outcome of the first toss, then she must revise her degree of belief about the branch weights via conditionalization. That is, an agent who sees heads on the first toss will have degree of belief 2/3 that the coin is biased towards heads. Of course, there will be branches on which our agent's successors are misled, and *decrease* their degrees of belief in the true hypothesis about branch weights. But on a higher weight of branches the agent's successors will have their belief-states improved.[6]

[6]This can be made precise: it can be shown, via the argument of Greaves and Wallace (2006), that, on any reasonable way of measuring the epistemic value of a belief-state, updating by conditionalization maximizes expected epistemic value. This epistemic-utility argument is complementary to the intertemporal-consistency defence of conditionalization given in the Appendix.

It is a consequence of updating beliefs about branch weights by conditionalization that agents on all branches take the results of experiments to favour hypotheses that afford their own branches high weight, and so boost their degrees of belief in such hypotheses and lower their degrees of belief in hypotheses that afford their own branches low weight. The copy of our agent on each branch ends up believing that the set of branches that share the outcome that she has seen has high weight. Some of them will be mistaken, of course. But there will be a higher total weight of agents who have had their beliefs about branch weights altered in the direction of the truth, than of those who have been misled.

5 Objections and replies

As our presentation above has tried to emphasize, there is a pervasive structural analogy between chance theories and branching-universe theories (and between chances and branch weights, and between possible worlds and branches). Correspondingly, many of the objections that might be raised against the proposed account of decision-making and/or belief-updating in the face of branching have equally compelling (or uncompelling) analogs in the chance case. This is important: we claim only that the Everett interpretation is *no worse off* than any other theory vis-a-vis the philosophy of probability, so any objection that applies equally to both cases will be irrelevant to the present project.

Before considering particular objections in any depth, we therefore summarize the analogy that we see between the two cases. It will be helpful to keep this analogy in mind in the discussion that follows because, if there is to be a branching-specific objection, it must take the form of a claim that the analogy presented here is incomplete in some *relevant* respect; in every case, our replies will claim that it is not.

<table>
<tr><th>Chance setup (gamble)</th><th>Branch setup (bramble)</th></tr>
<tr><td colspan="2">Preferences between wagers go as maximizing expected utility, which is an average of utilities across</td></tr>
<tr><td>alternate possible outcomes</td><td>all branches</td></tr>
<tr><td colspan="2">weighted by an α-function. We call this α-function</td></tr>
<tr><td>a credence function.</td><td>a quasi-credence function.[7]</td></tr>
<tr><td colspan="2">For an exchangeable sequence of experiments, the agent's α-function can be represented as a mixture of extremal exchangeable functions α_λ. The agent acts as if she believes that one of these extremal functions is objectively the best one to base decisions on (although, in general, she is not sure which), and her α-function is an epistemically weighted average of them. That is, she acts as if the α_λ's are candidates for being</td></tr>
<tr><td>objective chance distributions,</td><td>objective branch weights,</td></tr>
<tr><td colspan="2">and the weighting (μ) of these that yields her α-function reflects her credences about which vector λ gives the right one. Updating by conditionalizing on results of experiments in the exchangeable sequence permits her to refine her credences about which α_λ is objectively best. Part of the content of quantum mechanics is the claim (which either is derivable from the non-probabilistic part of the theory, or is an independent postulate of the theory), that</td></tr>
<tr><td>chance = $|\text{amplitude}|^2$.</td><td>branch weight = $|\text{amplitude}|^2$.</td></tr>
<tr><td colspan="2">Call this the Born Rule. Note that the Born Rule is a substantive claim: left and right side of this equation have independent meanings (the left implicitly defined by decision theory, the right by quantum mechanics). Moreover, it is an empirically testable claim. Conditionalizing on results of experiments in the exchangeable sequence will cause the agent's credences about the values of</td></tr>
<tr><td>chances</td><td>branch weights</td></tr>
<tr><td colspan="2">to become peaked about the observed relative frequency. If the observed relative frequency is close to the value calculated from the Born Rule, it will raise credence that quantum mechanics is correct; if it is far from the Born Rule value, it will lower credence that quantum mechanics is correct.</td></tr>
</table>

[7] In Greaves (2004) and Greaves (2007a), the term 'caring measure' was also used. It was applied to the measures over branches that lie to the future of a given branch in a given multiverse that one obtains by conditionalising the agent's quasi-credence function (that is, her α-function) on the self-locating proposition that she is currently on the branch in question in the multiverse in question. There is thus a 'caring measure' that coincides with the quasi-credence function, and gives the agent's betting quotients, in the special case (and only in that case) in which the agent is sure which multiverse is actual and which branch in that multiverse she is on. (This comment is included only to clarify the

<table>
<tr><th>Chance setup (gamble)</th><th>Branch setup (bramble)</th></tr>
<tr><td>There will certainly be branches on which</td><td>There are possible worlds in which</td></tr>
<tr><td colspan="2">anomalous statistics occur. A frequentist analysis of</td></tr>
<tr><td>chance</td><td>branch weight</td></tr>
<tr><td colspan="2">is untenable: one cannot hold that</td></tr>
<tr><td>‘the chance of E is x’</td><td>‘the branch weight of E is x’</td></tr>
<tr><td colspan="2">just means ‘the long-run frequency of E will be x’ because, for any E and x,</td></tr>
<tr><td>it is possible that</td><td>there will be some branches on which</td></tr>
<tr><td colspan="2">the long-run frequency of E is not x. Relative frequencies are connected only evidentially with</td></tr>
<tr><td>chances.
In anomalous-statistics possible worlds,</td><td>branch weights.
On anomalous-statistics branches,</td></tr>
<tr><td colspan="2">agents are misled: they rationally lower their credence in the theory that is in fact true. Still,</td></tr>
<tr><td>the possible worlds in which this occurs have a low total chance.</td><td>the branches on which this occurs have a low total weight.</td></tr>
<tr><td colspan="2">There is no available updating policy that guarantees that agents raise their credence in the true theory</td></tr>
<tr><td>in every possible world.</td><td>on every branch.</td></tr>
<tr><td colspan="2">It therefore makes sense that the conditionalization strategy recommended is the optimal one:</td></tr>
<tr><td>low-chance</td><td>low-weight</td></tr>
<tr><td colspan="2">events don’t count for much in the evaluation of wagers.</td></tr>
</table>

We now consider and reply to eight foreseen objections to the account proposed in the present paper.

Objection 1: Branch weights are not probabilities. Reply: we do not claim that they are. The claim is, rather, that, given reasonable constraints on an epistemic agent’s preference between wagers, she will act as if she believes that there are physical branch weights, analogous to physical chances, that can be estimated empirically in the same way that chances are, and that observation of events to which a theory assigns high branch weight boosts

relationship between the three papers in question; the concept of ‘caring measure’ plays no special role in the present paper.)

rational credence in a branching theory in the same way that observation of events to which a theory assigns high chance boosts rational credence in a chance theory.

Objection 2: The decision-theoretic account is all about the behavior of rational agents; this is (surely) irrelevant to matters of physics, and so cannot supply the Everett interpretation with an acceptable account of *physical* probability.

> We do not accept that the behavior of a rational decision maker should play a role in modeling physical systems. (Gill, 2005)

The reply to this has two parts. The first is that there is a clear sense in which, in order to model a physical system, one does not need to invoke considerations of rationality, and that this remains true in Everettian quantum mechanics. The second (and deeper) point is that—the first point notwithstanding—considerations of rationality have always played a role, and indeed must play a role, in the *confirmation* of physical theories, so it is no objection to the approach outlined above that it brings rationality considerations into the discussion of the confirmation of Everettian quantum mechanics. Let us explicate each of these two points in turn.

First, the sense in which the modelling of physical systems is silent on issues of rationality. The point here is perfectly straightforward. According to quantum theory (Everettian or otherwise), one models a physical system by ascribing to it a quantum state—a vector in, or density operator on, some Hilbert space. On the Everettian account of measurement, after a measurement there will exist a multiplicity of branches; the quantum state of the universe will be a superposition of the states of these branches, with some particular set of complex coefficients (amplitudes). Here we have, in outline, a physically complete account of the situation before and after measurement, and nothing has been said about rationality.

Now let us move on to the second point: that considerations of rationality *must* be relevant to theory confirmation. The point can be seen abstractly as follows. The question under consideration — when we are talking about theory confirmation—is that of *which physical theories it is rational to believe* (or have significant degree of belief in, or have significant degree of belief in the approximate truth of, etc.), given the evidence we in fact have. This is a question about a *relation* between physical theories on the one hand, and rationality on the other. It should then be of no surprise that, in answering the question, we need to consider the theory of rationality, as well as our various candidate theories of physics.

The point can be made more vivid by considering a more concrete case. Let us put the issue of branching-universe theories aside for the moment. Suppose that we have a physical theory, call it T, that is irreducibly stochastic. (T can be thought of as, for example, a dynamical collapse theory along the lines proposed by GRW *et al.*) Consider some fixed experimental setup, A. Suppose fixed the way that A is to be modelled in terms of T (including initial conditions). Then, according to T, there are a number of possible outcomes for the experiment A: $s_1, \ldots, s_n$. T also assigns chances to the various possible outcomes: $p_1, \ldots, p_n$ for $s_1, \ldots, s_n$ respectively. But now suppose that these so-called 'chances' are unrelated to considerations of rationality. In particular, suppose that there is no rationality constraint to the effect that the experimenter, insofar as she believes T, should bet at odds given by $p_1, \ldots, p_n$ on the outcome of the experiment; and that there is no rationality constraint with the consequence that, if in a long run of repetitions of the experiment she observes relative frequencies that approximately match the single-case chances predicted by T, and that no other available theory has this so-called 'virtue', then she should increase her degree of confidence in the theory over its rivals. Under these suppositions, *the 'chances' ascribed by the theory would have become altogether idle*: for all practical and theoretical purposes, we would be no better off than if our theory merely said that such-and-such a range of outcomes was *possible*, and ascribed the various possibilities no chances at all. In particular, the evidential connection between theoretical single-case chances (on the one hand) and observed relative frequencies (on the other) can be made to reappear only by admitting the connection between chances and rational belief revision.

Why is this point often missed? In our view, the explanation is the prevalence of (a) a frequentist analysis of chance and (b) a falsificationist account of confirmation—both of which accounts are importantly defective. On the combination of these two (defective) accounts, one reasons as follows. First, one takes it that one knows perfectly well what to do with predictions of the form 'the chance of E is x', without touching on issues of rationality: such predictions *just mean* (according to frequentism) that in a long run of repetitions of the experiment, the relative frequency of E will be approximately x. Second, one notes (as a consequence) that if the observed relative frequency deviates significantly from the theory's single-case chance, then something has happened that the theory predicted would not happen, and hence has been falsified; if, on the other hand, there is approximate agreement (and, perhaps, the prediction was a 'risky' one), then the theory has been confirmed (or, perhaps, 'corroborated').

The deficiencies of frequentism and falsificationism are well known. To repeat: a probabilistic theory does *not* predict, *categorically*, that the observed relative-frequency *will* approximately match the theoretical single-case chance; what it predicts that (in a sufficiently long run of experiments) this matching will be observed *with probability close to one*. So the first assertion in the above frequentist-deductivist account is false. And when one replaces 'the theory predicts that the observed relative frequency will approximately match the theoretical single-case chance' with 'the theory ascribes probability close to one to the proposition that the observed relative frequency will approximately match the theoretical single-case chance', the second step in the above account develops a glaring hole: if the observed relative frequency deviates significantly from the theory's single-case chance, then something has happened that the theory *ascribed low probability to*, but this is perfectly consistent with the theory's being true, so the theory has not been falsified.

The would-be deductivist is then tempted to patch up the account with a principle to the effect that, if something happens that the theory deemed *sufficiently improbable*—say, to which the theory ascribed probability less than some threshold p_{thresh}—then the theory is to be regarded as effectively falsified. But this patching-up will not work either: for every way the observations could turn out (including relative frequencies that approximately match the theoretical single-case chances), there will be some description under which 'those observations' were astronomically improbable (such as the particular ordered sequence of outcomes observed).

To escape from this quagmire, one must move to something more closely resembling a Bayesian account of theory confirmation. But then, if one really wants to be precise about the details, one is up to one's elbows in rationality constraints—on belief-updating, and on the connection between conditional credences and chances (the Principal Principle). The account we have given is just the extension to the branching case of this standard Bayesian account. To be sure, one can, for the purposes of most discussions of physical theory, avoid *explicit* discussion of rationality. One can simply help oneself to a particular consequence of the Bayesian theory: the principle (CC) stated in section 2. The same thing can be done in the Everettian case: one can simply help oneself to the principle (CW) stated above. This, too, obviates the need to write several paragraphs on decision theory before drawing evidential conclusions from a laboratory experiment. But it is a myth that the foundations of these confirmation-theoretic principles are independent of the theory of rational belief and decision. Our task in this paper has been to provide the foundation for the principle (CW): it is only for this reason that our discussion has been more explicitly rationality-theoretic than that found in the average physics text.

Objection 3: The decision-theoretic account presented here shows that agents must attach *some* decision-theoretic weights to branches, but it does not show that these weights must equal those given by the Born rule. There is a sense in which this is correct, and a sense in which it is not.

The observation made by the 'objection' is correct in the sense that we have not supplied a 'derivation of the Born rule' from the pre-existing part of the theory. That is, we have not supplied an *a priori* proof that betting quotients for outcomes, conditional on the truth of Everettian quantum mechanics *stripped* of any explicit postulate about the relationship of *branch weights* to (say) the amplitude-mod-squared measure over branches, must, on pain of irrationality, be those given by the Born rule. That is, we have not made the claim that is made by Deutsch (1999) and Wallace (2003, 2007). However, as we will now explain, we do not take this to be ground for any objection to our account.

The status of derivations of Born-rule weights within Everettian quantum mechanics is (at least *prima facie*) similar to the status of Gleason's theorem and related results[8] concerning probability in quantum mechanics. They show that certain assumptions lead to Born rule probabilities (or weights). The assumptions used as premises in such proofs are not beyond question. At most, such proofs show that Born-rule chances/ignorance-probabilities/ branch weights are the only ones that fit naturally with, or, perhaps, are definable in terms of, the *existing* structure of quantum mechanics. It is an open question what the significance of this is: whether, chance/ignorance-probability/weight predictions *should* be thus definable in terms of the structure that is already present in the theory prior to the introduction of chances/ignorance-probabilities/weights. Further, it is conceivable that the answer to this question could turn out to be different in a chance or an ignorance-probability theory than in a branching-universe theory. For example, Wallace (2007, section 6) can be understood as arguing that in the branching case decision-theoretic branch weights must be definable in terms of the structure of the theory, but that the analogous claim for chances or ignorance probabilities is not true; meanwhile, the existence of the subject of nonequilibrium statistical mechanics, of the work of Valentini *et al* on Bohmian mechanics 'out of quantum equilibrium', (see *e.g.*, Pearle and Valentini (2006)) show (for whatever this is worth) that the principle that probabilities be definable in terms of pre-existing structure is in fact flouted

[8]A nice recent example is the Zurek's derivation of the Born rule from envariance Zurek (2005). Barnum (2003) has shown how to turn this proof into one that takes no-signalling as its main premise.

by some (non-branching) theories taken seriously by working scientists.

Perhaps probabilities and/or weights must be derivable from pre-existing structure; perhaps this issue plays out differently in branching and non-branching theories; perhaps not. We have no further comment to offer on these issues. Fortunately, such issues are irrelevant to our central claim. If chances/ignorance-probabilities/branch weights must be definable in terms of the existing structure of the theory, then the Born rule *seems* to be the only option, in either a branching or a non-branching version of quantum mechanics. In that case, had we consistently observed non-Born frequencies, we would have been compelled to abandon quantum theory altogether. If (on the other hand) there is no requirement that weights be definable in terms of existing structure, then two versions of Everettian quantum mechanics that agree on everything but the branch weights but ascribe non-Born weights to branches are, for the purposes of theory confirmation, distinct theories, just as Bohmian mechanics in and out of quantum equilibrium are distinct theories. In this case, had we observed non-Born frequencies, we would have had more latitude; it is the package as a whole, branches (or possible histories) plus branch weights (or chances, or ignorance-probabilities) that is confirmed or disconfirmed; it might have been open to us to retain a core of quantum theory, but to adopt a different rule for the chances/ignorance-probabilities/branch weights.

As things have turned out empirically, however, this is all largely irrelevant: we have observed Born frequencies and so, whether or not there are other coherent theories out there that otherwise agree with Everettian quantum mechanics but postulate non-Born branch weights, the theory with Born-rule branch weights has been empirically confirmed (and any candidate theories with non-Born branch weights have been disconfirmed). This is the sense in which the objection is *in*correct: our account *does* have the consequence that—whether or not there exists a satisfactory 'derivation of the Born rule' from the pre-existing part of the theory—rational agents who observe long runs of Born-rule frequencies will increase their degree of belief that the weights of future branches are those given by the Born rule.

Objection 4: There are branches on which non-quantum statistics are observed. Hemmo and Pitowsky write:

> Even for agents like us, who observed up to now finite sequences which *a posteriori* seem to conform to the quantum probability [*i.e.* the Born rule], adopting the quantum probability as our subjective probability for future action is completely arbitrary,

> since there are future copies of us who are bound to observe frequencies that do not match the quantum probabilities (Hemmo and Pitowsky, 2007, p. 348).

The inference from the existence of branches to the arbitrariness of adopting Born-rule probabilities as guides to future choice requires some explanation. The argument seems to be something like this. On the ordinary account, evidence from past relative frequencies provide grounds for believing, if not with certainty then at least with high degree of belief, that future relative frequencies will be similar. But, on an Everettian account, there are no grounds for such belief, and we are in fact *certain* that relative frequencies will deviate arbitrarily far from Born-rule weights on *some* future branches. In the absence of an account on which observation of past frequencies is evidence that it is better, in some sense, to adopt Born-rule probabilities as guides to future actions, these past observations are irrelevant to future action.

We claim to have supplied such an account. Theorem 2 shows that updating beliefs about branch weights by conditionalizing on observed data is preferred to any other strategy. In worlds like ours, provided only that the agent regards the sequences of experiments in question as repeatable, this leads to beliefs that the Born-rule branch weights are at least approximately correct. It is, therefore, not arbitrary.

If our agent has priors on which a sequence of experiments is exchangeable, and if these are non-dogmatic, then she will treat past experience as relevant to future action. Of course, an agent might have priors that are such that the result of one experiment is never relevant to that of another, and so be unable to learn from experience which betting strategies are better than which other. But she would be an agent who could not do science. We are not aware of any reason for thinking that the sort of assumption that entail learning from experience are any less reasonable in the branching than in the non-branching case.

Objection 5: According to the Everett interpretation, what the observer learns when she observes a measurement outcome is only self-locating information. This cannot possibly be relevant to theory confirmation. The idea here is as follows. Consider an agent who is about to perform some quantum measurement with n possible outcomes $O_1, \ldots, O_n$. Conditional on the proposition that the Everett interpretation is true, this agent is *certain*, for each value of i from 1 to n, that there will be some future branch on which O_i occurs, and some future copy of herself on that branch. The measurement is then performed. A later copy of our agent looks at the apparatus in her lab, and observes that, on her branch, some particular outcome O_j occurred. Then (the thought runs) *conditional on the truth of Everettian quantum mechanics, the information she has acquired is purely self-locating* — she knew all along (conditional on the truth of Everettian quantum mechanics) that there would be such a copy of herself, and now she has merely observed that indeed there is. Therefore (the objection continues), she cannot possibly have learnt anything that is evidentially relevant to the truth of Everettian quantum mechanics. (The thought is related to that raised in objection 4.)

Let us put aside the awkwardness ('knew conditional on the truth of Everettian quantum mechanics', etc.) required to state the sense in which the information is 'purely self-locating'. The key mistake on which the above objection rests is the idea that information that is 'purely self-locating', in the sense that it does not rule out any possible worlds, is necessarily also evidentially irrelevant to *de dicto* propositions (i.e., that it cannot, under rational belief-updating, result in the redistribution of credences between possible worlds).

Such a principle cannot be sacrosanct; there are in any case many known counterexamples, independent of the Everett interpretation. Consider, for example, the prisoner in a lighted cell, who knows that it is six o'clock in the evening and that the light in her cell will be switched off at midnight iff she is to be hanged at dawn. Some significant amount of time passes, and the light stays on; the prisoner rationally becomes more confident that she will live another 24 hours. But nothing that she has learnt *rules out* the possibility that she will be hanged at dawn: she remains uncertain as to whether or not midnight has really passed. For a second (more familiar, but also more controversial) example, consider Sleeping Beauty: the two most common analyses of Beauty's case, the 'thirder' (Elga, 2000) and 'halfer' (Lewis, 2001) analyses, agree that on learning that it's Monday, Beauty acquires evidence that the coin landed heads, despite the fact that her being awake on Monday is *consistent* both with Heads and with Tails.

Our account is one according to which this (anyway non-sacrosanct) principle is routinely violated: information about the outcomes of experiments (in the possible world in which, and/or on the branch on which, the agent is now located) is a type of information that, even in the highly idealized cases in which it becomes purely self-locating (i.e. cases in which the agent is certain that some branching-universe theory is correct), *is* evidentially relevant to *de dicto* propositions. Furthermore, we are aware of no well-motivated alternative account of belief-updating that renders it evidentially *ir*relevant. (The methodological point implicit in this reply is that it is often more reliable first to work out which global belief-updating strategies are candidates for rational status, and afterwards to draw conclusions about which sorts of information can be evidentially relevant to which sorts of propositions, than vice versa.)

Objection 6: Decision-making is incompatible with deterministic physics. This issue is an old one. If the underlying physical dynamics is deterministic, then the decision our agent is going to make is already determined by the present state of the universe, together with the dynamical laws. It is an illusion, according to this objection, that she has any decision at all to make.

In reply, it should first be mentioned that this does not differentially affect our account, but applies equally well to any deterministic physical theory. Nor does a move to an indeterministic physics help; making my actions partly a matter of chance does not address the concerns behind this objection.

Fortunately, we do not have to consider the age-old problem of freedom of the will here. Our axioms concern *rational* preferences between wagers. It makes sense to have such preferences, and to evaluate them as rational or irrational, independently of questions of our ability to act on the basis of such evaluations. We frequently evaluate our own actions (often, negatively) in cases in which, due to weakness of the will, we are unable to act in the way that we judge to be best.

Objection 7: Preferences between wagers is nonsensical in a branching universe. On the branching-universe view, *all* payoffs corresponding to non-null outcomes are actually paid to agents on the corresponding branches. This is certainly a departure from the usual way of thinking about wagers. Some readers may find themselves at sea when contemplating such a scenario, and it may seem that we have no clear ideas about what preferences between branching wagers might be reasonable.

We claim that the situation is not so grim. For one thing, there do seem to be some clearly defensible principles regarding rational preferences between branching wagers. A wager that pays a desirable payoff on all branches is surely preferred to a wager that pays nothing on all branches. If, on every otucome, the payoff paid by $\mathbf{f}$ is at least as desirable as that paid by $\mathbf{g}$, then $\mathbf{f}$ is at least as desirable as $\mathbf{g}$.

If we accept that preferences between wagers makes sense in the branching case, and accept also that there are principles that reasonable preferences between branching wagers ought to satisfy, then the question still arises whether the axioms we have laid down are reasonable constraints on rational preference in the branching case.

Objection 8: The decision-theoretic axioms are *not* as defensible in the branching as in the non-branching case. This is the most serious objection. If the axioms are accepted for preferences between wagers on a branching scenario, then confirmation of theories that posit branch weights proceeds in a manner entirely parallel to theories that posit chances.

One occasionally comes across the following idea: since decision-making conditional on the assumption that the Everett interpretation is true is decision-making under conditions of certainty, 'the' decision theory for such decision scenarios is trivial (meaning: it consists merely of the requirement that preferences be total and transitive, that is, our axiom P1).[9]

If intended as an objection to the account defended in this paper, this point would beg the question entirely. One can, of course, write down both trivial and nontrivial decision theories, both for decision making in the face of indeterminism and for decision making in the face of branching. The fact that decision making in the face of branching had not been seriously considered (and hence no nontrivial decision theories for that case advocated) prior to 1999 is irrelevant; the question is which decision theories are *reasonable*. Our claim is that the nontrivial decision theory we have outlined is no less reasonable in application to the branching case than in its long-accepted application to the structurally identical indeterministic case. A non-question-begging objection in this area must give a *reason* for thinking that, structural identity notwithstanding, the decision-theoretic axioms we have discussed, while reasonable in the indeterministic case, ought not to be applied to branch setups in the way that we have advocated.

Such a reason must involve a difference between chance setups and branch setups. The most obvious difference is that, on a branching scenario, all outcomes actually occur, whereas, in the non-branching scenario, only one outcome is actual. Moreover, if a given payoff is paid only on a class of outcomes with low chance, our agent can be reasonably certain that that payoff will not be the one that is paid. In the branching scenario, the corresponding payoff is sure to be paid, albeit on a class of outcomes with low weight.

Why might this difference be relevant? We will not explore the full range of possible reasons here; we discuss only the one that, in our own opinion, poses the most serious *prima facie* problem for our account. It is this. The fact that all outcomes actually occur supplies a sense, *possibly* relevant, in which preferences between wagers on branch setups ('brambles', as, following Barry Loewer, we called them in the above table) are analogous to questions

[9] Mostly this suggestion has been made in conversation; however, see also Wallace (2002, section 3.2).

of distributive justice.[10] Now, the representation we obtain from our axioms P1–P6 is one in which alternate distributions of payoffs are judged according to a weighted mean of utilities on all branches. Though some, namely utilitarians, accept that judgments regarding distributive justice, no less than questions of rational decision under uncertainty, are to be addressed in this way, there are of course dissenters from such a view. Rawls, for example, argues for a maximin rule, which seeks to maximize the well-being of those that are worst off (Rawls, 1999). Someone who accepts such a view for questions of distributive justice could still think that rational strategies for prudent decision-making under uncertainty conform to the axioms. If she regards preferences among brambles as relevantly similar to questions of distributive justice rather than to preferences among gambles, she will then accept our axioms for chance setups, but be wary of a rule for *branch* setups that ranks wagers according to a weighted mean of payoffs on branches.

Two replies can be made to this objection. The first is that even if preferences among brambles is relevantly similar to questions of distributive justice, still the representation theorem discussed in section 3 will do the epistemic work we have claimed it can do in the branching context. The second is that it is at least far from obvious that the similarity in question *is* relevant. We will set out these two replies in turn.

The first reply runs as follows. We concede (for the sake of argument) that preferences among brambles are relevantly similar to questions of distributive justice, but we claim that, *for a suitably restricted class of decisions*, even questions of distributive justice are suitably treated using the weighted-average formula. The point is the following: all that is required, for the representation theorem to go through, is that there be at least two payoffs, one of which is strictly preferred to the other. This means that we do not need to make the strong claim that our axioms apply even when some of the brambles among which the agent is choosing assign a *terrible* outcome to some branches, or in general when the utility differences between branches are large. (This is relevant because part of the intuition underlying the objection is that there are some things that we ought not do to *anyone*, no matter how great the benefit to others might be.) Suppose, then, that we restrict our attention to preferences between wagers involving only the payoffs, a: receive one chocolate donut, and b: receive two chocolate donuts, with b strictly preferred to a. The trivial wager $I(b)$ is strictly preferred to the trivial wager $I(a)$. All other wagers in this restricted class are ties on the maximin rule, since they share the same worst outcome. It does not seem reasonable to simply be indifferent between all such wagers; if $\mathbf{f}$ and $\mathbf{g}$ coincide except

[10] See Huw Price's contribution to this volume.

on a non-null class F of outcomes, on which $\mathbf{f}$ awards a and $\mathbf{g}$ awards b, then, surely, $\mathbf{g}$ is to be strictly preferred to $\mathbf{f}$. We claim that axioms P1-P6 are reasonable ones for breaking maximin ties within this restricted class of wagers. And, if this is accepted, then we still get a representation on which values of wagers are represented as a weighted mean of utilities on branches. For the confirmation-theoretic purposes of this paper, we do need to claim that a rational agent should always have the same preferences when faced a bramble or with a corresponding gamble. Perhaps a case can be made for this strong claim; but we need not take the analogy so far.

The second reply, rather than conceding the objector's point and arguing that it is not damaging, challenges the point itself, as follows. It is far from obvious that Everettian decision-making is *relevantly* analagous to distrIbutive justice, rather than to decision-making under classical uncertainty, in cases (if any such there be) in which the correct decision procedures for the latter two situations diverge. To be sure, as we noted above (and as our objector emphasizes), brambles and distributive justice problems share the attribute that all candidate reward-recipients are actual. But, since any two scenarios are similar in some respects and dissimilar in others, the existence of *some* criterion effecting this grouping is trivial. There are, of course, many other (more or less natural) criteria that would group brambles and gambles together while excluding distributive justice problems, and still others that would group gambles and distributive justice problems together while excluding brambles. (An example of the former type of criterion is: are the candidate reward recipients future copies of the decision-making agent?[11] An example of the latter is: was the scenario in question discussed prior to 1950?)

What is required, in order to assess the relevance or otherwise of appeals to distributive justice, is a careful exploration of precisely which differences between scenarios of classical uncertainty and those of distributive justice are responsible for the divergence in recommended policies; only once we have such an explanation can we know to which the Everettian case is relevantly analogous. This project has not (to our knowledge) been carried out, and lies beyond the scope of the present paper. But in the meantime, it is at least plausible that the fairness-based intuitions that motivate deviations from maximization of expected utility in the case of distributive justice are grounded in issues of trust and power dynamics, present in a complex community of distinct and interacting agents but absent in the case of 'brambles',

[11] Following a panel discussion at the Perimeter Institute conference during which one of us raised this point, Simon Saunders suggested an alternative that is probably more to the point: Is there any interaction between the recipients?

and have nothing at all to do with the mere fact that all candidate recipients are actual. (When deciding whether to increase contributions to one's pension fund or blow the extra money on an expensive holiday next year, one doesn't worry about whether or not one's allocation of resources between one's next-year self and one's older self is *fair*, despite the fact that both are [timelessly] actual. It is very interesting to ask why not, and we do not know the answer; but the datum is clear.) It is thus at least plausible that the analogy to cases of distributive justice is irrelevant.

Let us turn to P7. This axiom *seems* non-controversial: unless our agent has cause to re-evaluate her earlier judgments about preferences between wagers, she should continue to employ the updating strategy that, on her initial preferences, she deemed the best. This, as we have seen, is equivalent to updating by conditionalization. The following objection, however, can be raised to P7 in the branching case.[12]

The account defended in this paper has the post-branching agent adopting the updating strategy ranked highest by the pre-branching agent. But (the objection runs) our post-branching agent *knows* that, if the Everett interpretation is true, then her interests now are not the same as the interests of her pre-branching self — the latter's interest was to maximize average utility across branches according to the measure of importance of those branches, whereas the former's interest is to maximize utility on whichever branch she is in fact now on. And (the objection continues) intertemporal consistency criteria — such as P7 — can have the status of rationality constraints only if the agent-stages concerned believe that they have the same interests as their temporal counterparts. Therefore (the objection concludes), P7 is a rationality constraint for chance setups but not for branch setups.

Before replying, let us illustrate the objection by elaborating on the sort of example that it suggests. We imagine a situation in which, prior to a sequence of two coin flips, our agent (call her Alice_0) weighs options and decides whether, in her estimation, her successors on each of the post-flip branches should prefer wager **f** or wager **g** on the outcome of the second flip. Now let Alice_1 be a successor on one of the branches after the first coin flip, but suppose that Alice_1 has not yet learned the outcome of this first coin toss. We imagine that Alice_1 opts not to take the advice of Alice_0, on the grounds that her interests are different. However, since the situation is meant to be one of the sort for which updating by conditionalization would be required in the non-branching case, we must stipulate that Alice_1 still endorses Alice_0's judgments as appropriate for Alice_0's situation. (Recall

[12]We are grateful to Tim Maudlin for raising a similar objection to a predecessor of the position defended in this paper, and for extensive discussion.

that P7 restricts the intertemporal consistency requirement to 'pure learning experiences'.) Alice$_1$'s reasoning must then be: "Alice$_0$, in formulating her advice, was concerned with maximizing mean benefit across all branches to *her* future. But I'm only concerned with myself and with my future branches. (And, by the way, I see no reason for my self-locating credences concerning which branch I am now on to bear any particular relationship to Alice$_0$'s estimates of the relative importance of branches.) Though Alice$_0$ was right, given her concerns, to recommend that I choose **f**, it would be better for me, with my concerns, to choose **g**."

We claim that Alice$_1$ should, rather, accept Alice$_0$'s advice on whether to take **f** or **g**. The reason is that there is a relevant sense in which the interests of Alice$_0$ and Alice$_1$ are the same: they both ultimately aim to maximize actual payoff averaged with respect to the actual branch weights. But given that they do not know the actual branch weights, each's preferences over wagers are given by maximizing average payoff with respect to their respective credences about the branch weights. Now, the actual branch weight of a given payoff on a wager on the second coin flip is the same downstream of Alice$_1$ as it was downstream of Alice$_0$, and Alice$_1$ knows this. And Alice$_1$ has gained no new information about the branch weights; she is, with respect to branch weights, in a 'pure learning situation' in which nothing has been learned. If she endorses Alice$_0$'s credences, she should therefore retain the same credences about branch weights, and hence the same preferences among wagers. Hence Alice$_1$ will endorse Alice$_0$'s recommendations about what she should do upon learning the result of the experiment. The subsequent learning of the outcome of the experiment involves no branching, merely a gain in knowledge, so there is no room for a supposed change of interest to alter her judgments about what she should do, upon learning which sort of branch she is on.

Someone might accept P1-P7 and nevertheless insist on a more egalitarian treatment of measurement outcomes, continuing to bet at even odds on the outcome of a coin toss, even in the face of a string of tosses in which heads predominate. After all, the argument goes, there will be a copy of me on the H branch, and a copy of me on the T branch; ought not I be fair, and treat both of these copies equally? This amounts to rejecting P8, which is meant to exclude dogmatism of this sort.

This is reminiscent of Laplaceanism, and the reply is similar. Counting branches is not as simple as that. After all, there are many ways in which the coin can land heads, and many ways in which it can land tails. A wager that makes the payoff depend only on heads or tails imposes a partition on the outcome space. There is no necessity, and no compelling reason, why this partition, rather than some other, must be treated so that copies of me in one class hold the same weight in my deliberations as copies of me in the other class. Just as, in applications of probability conceived in the ordinary way, Laplace's definition of probability must be supplemented by judgments of *which* classes of events are to be judged equipossible, so, too, would an egalitarian approach to preferences between branching wagers require a judgment of which partitions of an experiment's outcome space are to be afforded equal weight. What we suggest is the same for the branching case as for the non-branching case: let experience be your guide.

6 Conclusions

Everettian quantum mechanics ascribes weights to branches. We have outlined an account according to which rational agents use these weights as if they were chances in evaluating bets that may give different payoffs on different branches, and the occurrence of events to which the theory ascribed a weight higher than the average chance-or-weight ascribed by rival theories increases rational degree of belief in the Everettian theory. That is, on this account, branch weights play both the decision-theoretic and the confirmation-theoretic role that chances play. We have argued that this account is no less defensible than the structurally identical account according to which chances, in an indeterministic theory, have similar decision-theoretic and confirmation-theoretic relevance. It follows from the same decision-theoretic axioms, via the same representation theorems; and, we claim, the axioms are no less plausible under our suggested interpretation in branching contexts than they are under the familiar interpretation in non-branching contexts. If correct, this solves the *prima facie* evidential problem that the Everett interpretation seemed to face.

7 Acknowledgments

Would like to thank the organisers of and participants in two conferences at which this paper was presented — the 'Everett at 50' conference held in Oxford in July 2007, and the 'Many Worlds at 50' conference held at Perimeter Institute in September 2007 — and the participants in a seminar at the University of Western Ontario in June 2007. Special thanks are due to Sona Ghosh, Bill Harper, Barry Loewer, Tim Maudlin, Simon Saunders and David Wallace.

References

Albert, D. and B. Loewer (1988). Interpreting the many worlds interpretation. *Synthese 77*, 195–213.

Barnum, H. (2003). No-signalling-based version of Zurek's derivation of quantum probabilities: A note on "environment-assisted invariance, entanglement, and probabilities in quantum physics". quant-ph/0312150.

Bernoulli, J. (2006). *The Art of Conjecturing.* Baltimore: Johns Hopkins University Press.

Borges, J. L. (1941). *El Jardín de Senderos que se Bifurcan.* Buenos Aires: Sur.

Borges, J. L. (1962). The garden of forking paths. In D. E. Yates and J. E. Irby (Eds.), *Labyrinths: Selected Stories & Other Writings.* New York: New Directions.

De Finetti, B. (1937). La prevision: ses lois logiques, ses sources subjectives. *Annales de l'Institute Henri Poincaré 7*, 1–68. English translation in De Finetti (1980).

De Finetti, B. (1980). Foresight: its logical laws, its subjective sources. In H. E. Kyburg and H. E. Smokler (Eds.), *Studies in Subjective Probability.* Huntington, New York: Reobert E. Krieger Publisher Company. Translation of de De Finetti (1937).

Deutsch, D. (1999). Quantum theory of probability and decisions. *Proceedings of the Royal Society of London A455*, 3129–37. Preprint available online from www.arxiv.org.

Diaconis, P. and D. Freedman (1980). De Finetti's generalizations of exchangeability. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability, Volume II.* Berkeley and Los Angeles: University of Caifornia Press.

Diaconis, P., S. Holmes, and R. Montgomery (2007). Dynamical bias in the coin toss. *SIAM Review 49*, 211–235.

Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis 60*, 143–7.

Gill, R. (2005). On an argument of David Deutsch. In M. Schurmann and U. Franz (Eds.), *Quantum Probability and Infinite Dimensional Analysis: from Foundations to Applications. QP–PQ: Quantum Probability and White Noise Analysis, Volume 18.* World Scientific.

Greaves, H. (2004). Understanding Deutsch's probability in a deterministic multiverse. *Studies in History and Philosophy of Modern Physics 35*, 423–56.

Greaves, H. (2007a). On the Everettian epistemic problem. *Studies in History and Philosophy of Modern Physics 38*, 120–52.

Greaves, H. (2007b). Probability in the Everett interpretation. *Philosophy Compass 2*, 109–28.

Greaves, H. and D. Wallace (2006). Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind 115*, 607–32.

Hemmo, M. and I. Pitowsky (2007). Quantum probability and many worlds. *Studies in History and Philosophy of Modern Physics 38*, 333–350.

Huygens, C. (1660). *Van Rekeningh in Spelen van Gluck*. Amsterdam.

Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*, Volume 2, pp. 263–93. University of Caifornia Press.

Lewis, D. (2001). Sleeping beauty: reply to Elga. *Analysis 61*, 171–6.

Ore, O. (1960). Pascal and the invention of probability theory. *The American Mathematical Monthly 67*, 409–419.

Pearle, P. and A. Valentini (2006). Generalizations of quantum mechanics. In G. N. J.-P. Francoise and T. S. Tsun (Eds.), *Encyclopaedia of Mathematical Physics*. Elsevier.

Rawls, J. (1999). *A Theory of Justice*. Cambridge, MA: Harvard University Press. Revised edition.

Saunders, S. (2005). What is probability? In S. D. A. Elitzur and N. Kolenda (Eds.), *Quo vadis quantum mechanics?* New York: Springer.

Savage, L. J. (1972). *The Foundations of Statistics*. New York: Dover Publications, Inc.

Skyrms, B. (1984). *Pragmatics and Empiricism*. New Haven and London: Yale University Press.

Skyrms, B. (1994). Bayesian projectibility. In D. Stalker (Ed.), *Grue! : the new riddle of induction*, pp. 241–262. Chicago: Open Court.

Smith, D. (1959). *A Source Book in Mathematics*. New York: Dover Publications.

Wallace, D. (2002). Quantum probability and decision theory, revisited. Available online from http://philsci-archive.pitt.edu/archive/00000885/02/decarx.pdf.

Wallace, D. (2003). Everettian rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studies in History and Philosophy of Modern Physics 34*, 415–439.

Wallace, D. (2006). Epistemology quantized: Circumstances in which we should come to believe the Everett interpretation. *British Journal for the Philosophy of Science 57*, 655–89.

Wallace, D. (2007). Quantum probability from subjective likelihood: improving on Deutsch's proof of the probability rule. *Studies in History and Philosophy of Modern Physics 38*, 311–32.

Zurek, W. H. (2005). Probabilities from entanglement, Born's rule $p_k = |\psi_k|^2$ from envariance. *Physical Review A 71*, 052105.

Appendix

Experiments A, B are to be performed in succession. Our agent is to be informed which element of a partition $\{D_i^A \mid i = 1,...,n\}$ of S^A that the outcome of experiment A falls into, and then offered a choice between wagers $\mathbf{f} = [E_j^B \to f_j]$, $\mathbf{g} = [E_j^B \to g_j]$ on the outcome of B, where $\{E_j^B | j = 1,...,m\}$ is a partition of S^B. A *strategy* consists of a choice, for each i, of $\mathbf{f}$ or $\mathbf{g}$ as the preferred wager on B upon learning that the outcome of A is in D_i.

Define the strategy ϕ by

$$\phi_i = \begin{cases} \mathbf{f}, & \text{if } \sum_j \alpha(D_i^A \cap E_j^B)\, u(f_j) \geq \sum_j \alpha(D_i^A \cap E_j^B)\, u(g_j) \\ \mathbf{g}, & \text{if } \sum_j \alpha(D_i^A \cap E_j^B)\, u(f_j) < \sum_j \alpha(D_i^A \cap E_j^B)\, u(g_j) \end{cases}$$

Let $\bar{\phi}_i$ be the opposite strategy: if ϕ_i is $\mathbf{f}$, $\bar{\phi}_s$ is $\mathbf{g}$, and *vice versa.* ϕ's choices are such that, for each i,

$$\sum_j \alpha(D_i^A \cap E_j^B) \big(u(\phi_{ij}) - u(\bar{\phi}_{ij})\big) \geq 0.$$

We will say that ϕ strictly prefers ϕ_i to $\bar{\phi}_i$ iff

$$\sum_j \alpha(D_i^A \cap E_j^B) \big(u(\phi_{ij}) - u(\bar{\phi}_{ij})\big) > 0.$$

We will show that:

i). For any strategy ψ, $\psi \preceq \phi$.

ii). If, for some i, ϕ_i is strictly preferred to $\bar{\phi}_i$, then, for any strategy ψ that disagrees with ϕ's choice on i, $\psi \prec \phi$.

Let ψ be any strategy. $\phi \prec \psi$ iff

$$\sum_i \sum_j \alpha(D_i^A \cap E_j^B)\, u(\phi_{ij}) < \sum_i \sum_j \alpha(D_i^A \cap E_j^B)\, u(\psi_{ij}),$$

or,

$$\sum_i \sum_j \alpha(D_i^A \cap E_j^B)\ (u(\phi_{ij}) - u(\psi_{ij})) < 0.$$

There is no contribution to this sum from those i, if any, on which ϕ and ψ agree. When ϕ and ψ disagree, $\psi_i = \bar{\phi}_i$. For such i,

$$\sum_j \alpha(D_i^A \cap E_j^B)\, (u(\phi_{ij}) - u(\psi_{ij})) = \sum_j \alpha(D_i^A \cap E_j^B) \big(u(\phi_{ij}) - u(\bar{\phi}_{ij})\big) \geq 0,$$

and so,

$$\sum_i \sum_j \alpha(D_i^A \cap E_j^B)\,(u(\phi_{ij}) - u(\psi_{ij})) \geq 0,$$

or, $\psi \preceq \phi$. If, for any i,

$$\sum_j \alpha(D_i^A \cap E_j^B)\,\big(u(\phi_{ij}) - u(\bar{\phi}_{ij})\big) > 0$$

we have $\phi \prec \psi$ for any ψ with $\psi_i = \bar{\phi}_i$.

This gives us α_i^A up to an arbitrary scale factor. If we wish to normalize the updated α-function, so that $\alpha_i^A(S^B) = 1$, we have

$$\alpha_i^A(E_j^B) = \frac{\alpha(D_i^A \cap E_j^B)}{\alpha(D_i^A)}.$$