Proving the Principal Principle*

Wolfgang Schwarz
28 May 2014

Abstract. Any credible interpretation of objective chance should make sense of the connection between objective chance and rational degree of belief. Ideally, an account that identifies chance with some objective quantity X should be accompanied by a story that explains, from independently plausible assumptions, why X guides rational credence in the way captured by the Principal Principle. I provide this story for various Humean accounts of chance, including frequentist and best-systems accounts. Along the way, I also suggest a generalization of the Principal Principle that allows for dyadic and "indefinite" chances.

1 The Challenge

A noteworthy feature of physical probability, or chance, is its connection to rational belief. If you know that the coin you're about to toss has a 50% chance of landing heads, then you should give equal degree of belief to heads and tails. More generally, physical probability satisfies the following *Coordination condition*, where $Ch(A)=x$ is the proposition that the chance of A equals x.

> **Coordination (first pass)**
> Normally, if P is a rational prior credence function, A a proposition, and $P(Ch(A)=x) > 0$, then $P(A/Ch(A)=x) \approx x$.

This fact about prior credence often carries over to posterior credence because information about chance tends to screen off other information relevant to the outcome of a chance process. For example, the information that the previous toss of your coin came up heads should not affect your credence in the next outcome, if you also know that the chance of heads is 1/2. In general, objective chance satisfies the following *Resiliency condition* (compare [Skyrms 1980]).

> **Resiliency (first pass)**
> Normally, if P is a rational prior credence function, A a proposition, and $P(Ch(A)=x) > 0$, then there is a substantial range of propositions B such that $P(A/Ch(A)=x \wedge B) \approx P(A/Ch(A)=x)$.

* Thanks to John Barker, Kenny Easwaran, Alan Hájek, Marcus Hutter, Michael Smithson and Alastair Wilson for helpful comments and discussion.

Combining Coordination and Resiliency, we get a form of the *Principal Principle* (compare [Lewis 1980]).

> **Principal Principle (first pass)**
> Normally, if P is a rational prior credence function, A a proposition, and $P(Ch(A)=x) > 0$, then there is a substantial range of propositions B such that $P(A/Ch(A)=x \wedge B) \approx x$.

Lewis's formulation in [Lewis 1980] isn't restricted to normal cases, and states a strict rather than approximate equality. On Humean accounts of chance, this strict Principle is arguably false for certain "undermining" propositions A. Lewis's answer was to move to a more complicated "New Principle" (see [Lewis 1994]). For reasons that will become clear, I prefer to stay with a softened version of the old Principle.

Lewis also suggests that (i) chance should be indexed to a time t and world w, and that (ii) resiliency holds for all propositions B about the history of w up to t. (These propositions Lewis calls *admissible*.) I agree that chance should be indexed, although not necessarily to a time and a world. This requires some adjustments to the three principles above, which will be made in section 3. I do not follow Lewis in specifying a fixed domain of resiliency, mostly because I want my principles to cover not only forward-looking dynamical probabilities in fundamental physics, but also probabilities found in genetics, population dynamics or statistical mechanics. As [Skyrms 1980: 10–19] points out, every statistical theory comes with its own domain of resiliency, so there is little more we can say in full generality except that the domain includes a substantial range of propositions, including many propositions that one can easily come to know and that would otherwise be relevant to the proposition A under discussion.

I do not want to take side in the debate over whether probabilities in genetics, population dynamics or statistical mechanics deserve the name 'chance'. Whatever we call them, these probabilities display essentially the same connection to rational belief as dynamical probabilities in fundamental physics.

This connection puts a tight constraint on the interpretation of the relevant probabilities. After all, most probability functions do not satisfy Coordination and Resiliency. Lewis [1994] argued that any proposal to identify chance with some objective measure X must explain why X satisfies the Principal Principle. He conjectured that only Humean interpretations like his "best-systems" account can live up to this challenge. However, he never showed that best-systems probabilities satisfy the Principle. Every now and then, someone tries to do better. [Mellor 1971] and [Howson and Urbach 1993] try to derive the Principle for hypothetical frequentism, [Loewer 2004] and [Hoefer 2007] for versions of the best-systems analysis, [Deutsch 1999] and [Wallace 2012] for branch amplitudes in Everettian quantum mechanics. But many remain unconvinced by these attempts. Indeed, there is a growing consensus that the task is impossible – that no account of

chance, Humean or non-Humean, can explain the Principal Principle (see e.g. [Black 1998], [Strevens 1999], [Hall 2004]). In this note, I want to make another attempt at showing that Humean chances satisfy the Principal Principle.

Let me be clear about the goal. The point is not to justify the Principal Principle. The Principle may well be an analytic truth about chance and credence. It says is that chance plays a certain role. But when we consider different hypotheses about the nature of chance, we have to ask whether these candidates are apt to play the role. By analogy, consider the claim that Jack the Ripper (if he exists) committed such-and-such murders in the late 19th century. This may well be analytic. Nonetheless, when we consider different hypotheses about the identity of Jack the Ripper – that he is Lewis Carroll, or Barack Obama, or some metaphysically primitive entity – we have to ask whether there is any reason to believe that these candidates satisfy the Jack the Ripper role, i.e. that they committed the relevant murders.

The goal, then, is to show that on a given interpretation of 'chance', the Principal Principle follows from independently plausibly assumptions about rational belief – assumptions that do not depend on the interpretation of 'chance'. As we will see, this is not too hard if chance is interpreted as the characteristic function of the set of truths, or as relative frequency; the relevant mathematical theorems are mostly well-known and will be briefly reviewed in section 2. In section 3, I will suggest a generalisation of the Principal Principle that allows for indexed chance and does not require chance to deal with "single cases". Afterwards, I will return to the task of deriving the (now generalised) Principle, using first a frequentist and then a best-systems analysis of chance.

2 First steps

Let's begin with a simple case: the *fatalist* interpretation of chance. According to fatalism, only what in fact will happen has any chance of happening, and its chance is 1. The chance function Ch is the "omniscient" function that maps every true proposition to 1 and every false proposition to 0. (The domain may be somewhat restricted, to avoid liar-type paradoxes that arise if the "propositions" are sentences.) On this interpretation, $Ch(A)=1 \leftrightarrow A$ and $Ch(A)=0 \leftrightarrow \neg A$ are analytic, and given that analytic truths have probability 1, the Coordination condition $P(A \,/\, Ch(A)=x) \approx x$ reduces to $P(A \,/\, A) \approx 1$ and $P(A \,/\, \neg A) \approx 0$. Whenever $P(A) > 0$ and $P(\neg A) > 0$, these are trivial theorems of the probability calculus. Resiliency is also guaranteed, since $P(A \,/\, A \wedge B) = 1$ and $P(A \,/\, \neg A \wedge B) = 0$ are theorems for all B with $P(A \wedge B) > 0$ and $P(\neg A \wedge B) > 0$.

Fatalist chances are completely determined by the history of actual outcomes in the world. So here we have a Humean account that does entail the Principal Principle, without any additional assumptions about rational credence. The only problem is that fatalism

is not a plausible interpretation of the probabilities in statistical theories. (Incidentally, this shows that *pace* [Lewis 1980], the Principle does not exhaust our concept of chance.)

On the other hand, the fatalist result points towards a more general lesson. The fatalist chance function is a maximally informed credence function: it is the credence of an imaginary agent who knows absolutely everything about the world. On the best-systems account, the chance function can also be understood as very well-informed credence, corresponding to the beliefs of an imaginary agent who has access to all occurrent facts, but limited memory, so that she cannot store all these facts one by one. If the Principal Principle holds for maximally informed credence, does it also hold for lesser credence functions?

Here is a reason to think that it does. Let P be a rational prior credence function, and define Ch as P conditioned on the true answer to some question Q. Assume for simplicity that the possible answers to Q form a finite partition. Let A be any proposition and x any number for which $P(Ch(A)=x) > 0$. Let E_x be the disjunction of all possible answers E to Q with $P(A/E)=x$. (There is at least one such E, as otherwise $Ch(A)=x$ couldn't have probability > 0.) Since the answers are mutually exclusive, it follows that $P(A/E_x) = x$. Moreover, $P(Ch(A)=x \leftrightarrow E_x)=1$. So $P(A/Ch(A)=x) = P(A/E_x) = x$. That is, if Ch is a probability function that lies in between the rational prior credence P and the omniscient function in terms of what it knows about the events in a world, then Ch satisfied the Coordination condition.

This is encouraging, but it does not go far enough. For one thing, we also need to establish Resiliency. Moreover, most Humeans do not define chance as rational credence conditional on the answer to a certain question – although the pragmatist best-systems accounts of [Cohen and Callender 2009] and [Frigg and Hoefer 2010] come rather close.

The perhaps best known Humean account of chance is (finite) *frequentism*. Here chance is identified with relative frequency in a suitable sequence of events. For example, if 40% of the coin tosses in a certain sequence land heads, then the chance of heads, relative to that sequence, is 0.4.

Bruno de Finetti [1937] proved some important connections between rational belief and relative frequency. Consider a sequence of n coin tosses, with 2^n possible outcomes, represented by the vectors $\Omega = \{H, T\}^n$. Suppose the rational prior belief function P regards these outcomes as *exchangeable*, meaning that it assigns the same probability to any two outcomes $\omega, \omega' \in \Omega$ that agree in the number of heads and tails. Let $Ch(H) = r$ be the proposition that the relative frequency of heads in the sequence is r (for $r \in \{\frac{m}{n} : 0 \leq m \leq n\}$). Let H_i be the proposition that the ith outcome is heads (for $i \leq n$). Exchangeability then entails that

$$P(H_i/Ch(H)=r) = r.$$

Moreover, the longer the sequence, the more stable $P(H_i/Ch(H)=r)$ becomes under

conditioning on other outcomes:

$$P(H_i/Ch(H)=r) \approx P(H_i/Ch(H)=r \wedge H_j) \approx P(H_i/Ch(H)=r \wedge \neg H_j).$$

So exchangeability guarantees both Coordination and Resiliency, of a kind.

'Of a kind', because the chance function Ch and the credence function P apply to different objects. Credence is defined for token events, such as the first toss landing heads, H_1. The frequentist chance function Ch, on the other hand, only applies to event types, e.g. heads.

The correct response to this "type conflict" between chance and credence, I think, is to reformulate the Principles of section 1. This will be the topic of section 3. For now, let's choose a simpler response and extend the frequentist interpretation to events like H_1, by ignoring the reference class problem and letting every token event inherit its chance from the corresponding type. So the (extended) chance of H_1 is the relative frequency of H. To get a probability measure over the space of outcomes Ω, we also need to specify joint probabilities for different tosses. The most natural choice here is to treat them all as independent. Thus suppose r is the ratio of heads in the actual sequence. Then for any sequence $\omega \in \Omega$, we define

$$Ch_r(\omega) = r^h(1-r)^{n-h},$$

where h is the number of heads in ω.

In this way, every relative frequency r determines an extended frequentist chance function Ch_r over Ω. De Finetti famously showed that as n goes to infinity, any credence function P that regards the sequence as exchangeable converges to a mixture of such chance functions. In fact, P comes close to such a mixture as long as n is not very small. So if the sequence at issue is sufficiently long and we read $Ch=Ch_r$ as the hypothesis that Ch_r is the extended frequentist chance function (i.e., that r is the relative frequency of heads), then for all $A \subseteq \Omega$, the prior credence in A equals the expectation of the chance of A:

$$P(A) \approx \sum_r Ch_r(A)P(Ch=Ch_r).$$

This is a little weaker than the Coordination principle $P(A/Ch(A) = x) \approx x$. The principle itself fails, because Ch_r generally assigns positive probability to outcomes in which the relative frequency of heads is not r. For example, if A is the proposition that all tosses land heads, then $Ch(A)=x$ entails that the actual ratio of heads is $\sqrt[n]{x}$. For $x < 1$, this means that $Ch(A)=x$ is incompatible with A, so that $P(A/Ch(A)=x) = 0$.

This is an instance of the *undermining problem* for Humean accounts of chance. In response, Hall [1994] and Lewis [1994] argued that the Principal Principle should be

replaced by a more accurate "New Principle". In the present context, the "New" form of Coordination could be expressed as

$$P(A/Ch{=}Ch_r) = Ch_r(A/Ch{=}Ch_r).$$

On the frequentist account, this immediately follows from exchangeability: $Ch{=}Ch_r$ is the set of outcome sequences ω in which the relative frequency is r; Ch_r is uniform over this set, and so is P if it satisfies exchangeability.

We can also establish restricted, *ceteris paribus* versions of the old Principle. In particular, we have

$$P(H_1 \ldots H_k/Ch(H_1 \ldots H_k) = x) \approx x$$

as long as k is small compared to n. And we have Resiliency in the sense that for $j \neq i$,

$$P(H_i/Ch(H_i){=}r) \approx P(H_i/Ch(H_i){=}r \wedge H_j) \approx P(H_i/Ch(H_i){=}r \wedge T_j).$$

We have assumed that the rational prior credence P regards the relevant sequence as exchangeable. But this is problematic. For suppose you learn that the first 100 outcomes in some binary sequence about which you have no preconceptions are $101010 \ldots 10$. Intuitively, you should then be more confident that next two outcomes are 10 than that they are 01. This suggests that $101010 \ldots 1010$ had higher *prior* probability than $101010 \ldots 1001$, so your priors didn't regard the sequence as exchangeable. Indeed, frequentist definitions of chance are often restricted to *random* (i.e., irregular) sequences, rather than simple patterns like $101010 \ldots 10$. The information that the frequentist chance of heads in a given sequence is 0.5 therefore entails that the sequence is sufficiently random. But then $101010 \ldots 10$ is ruled out and the sequence isn't exchangeable relative to P conditioned on information about Ch_r.

Fortunately, all this doesn't really matter because we don't need full exchangeability. The proofs just mentioned still go through if we only require that the rational prior credence P gives equal probability to any two *sufficiently random* sequences with the same ratio of outcomes, assuming that swapping all 1s and 0s in a sequence does not affect whether it is sufficiently random. This restricted form of exchangeability is quite plausible: if all you know about a sequence is that it looks random and contains a certain ratio of 1s to 0s, you should arguably assign equal credence to random-looking sequences with that ratio.

A precise frequentist analysis would now have to define a suitable notion of randomness for sequences (without mentioning chance, of course); see [Eagle Forthcoming] for a survey of the usual candidates. Fortunately, the present results are valid on any sensible way of filling in these details. However, one plausible requirement may be worth pointing out: the relative frequencies in a "sufficiently random" sequence should not fluctuate much

between earlier and later parts of a sequence. This means that the sequence of states in a Markov chain are often not a suitable base for the identification of chance with relative frequency. A suitable base would here be the sequence of state transitions rather than the sequence of states. (Exchangeability with respect to such derived sequences is closely related to de Finetti's notion of *partial exchangeability*; see [Diaconis and Freedman 1980]).

In practice, when we toss a coin, or set up a Stern-Gerlach experiment, we rarely identify the occasion as the ith member of a certain sequence. Even if there is a privileged way of arranging the relevant events in a series, we rarely know at which position in the series we are. On the other hand, the probability that the *present* toss lands heads, given that the overall frequency is x, plausibly equals the average of the conditional probability that the ith toss lands heads, weighted by the probability that the present toss is the ith toss. So we also have Coordination and Resiliency for hypotheses about the *present* toss.

We may also exploit our ignorance of the present position to directly support the Principal Principle, without assuming any form of exchangeability. This was already pointed out by Bertrand Russell in [Russell 1948: 402ff.] – to my knowledge, the first explicit outline of a proof for the Principal Principle. Russell's argument goes as follows. Let C be a class of events, and D an arbitrary member of C. Given that r percent of Cs are H, what is your degree of belief that D is H? Answer: r. This is a consequence of the "arbitrariness" of the choice. Now when we toss a coin, we haven't literally chosen an arbitrary member of the relevant class – whatever that would mean. But our epistemic situation with respect to the outcome is typically just the same: if r percent of tosses land heads, and we have no further information about this particular toss, then it could just as well be any member of the class, so our credence in heads should be r.

I will return to this line of thought in sections 4 and 5, where I will also explain how the results established so far bear on the best-systems analysis. But first, I want to make the promised amendments to the Principal Principle.

3 Generalising the Principal Principle

As formulated in section 1, Coordination and Resiliency presuppose that the objects of chance are unrepeatable, single-case propositions like *heads on the 17th toss*, rather than repeatable event types like *heads*. As we've seen, this is incompatible with the most straightforward frequentist interpretation. It arguably also doesn't match the probability statements in actual statistical theories, which typically say that under such-and-such conditions C, outcome A has probability x. Here, C and A are naturally understood as event types – in other words, *properties* – that can be instantiated several times within a world. The statistical law that outcome A under condition C has probability x can be

understood as a "partial" counterpart of the strict law that all Cs are As. Like the strict law, the statistical law primarily states a relation between properties.

The principles from section 1 not only presuppose that chance applies to single-case propositions, they also neglect the relational character of chance: they don't take into account that chance might be relative to a condition, a time, or a reference class. This is not what we find in many statistical theories, and it contradicts several important accounts of chance: for Lewis, chance is relative to a time-world pair or history; for frequentists, chance is relative to an underlying sequence; for propensity theorists, chance is a measure of the causal tendency of a physical system in state C to produce outcome A.

The chance of A relative to C is a kind of conditional probability, but it need not satisfy standard laws for conditional probabilities. Without committing to a particular theory of chance, all we can say is that chance is a family of probability functions, indexed by a set of properties. More precisely, I will assume that a chance function can be modeled as a function Ch that maps every element C of some set Γ of mutually exclusive properties to a probability space $\langle \Omega_C, \mathcal{F}_C, Ch_C \rangle$, where Ω_C is again a set of mutually exclusive properties, $\mathcal{F}_C$ is a suitable algebra over Ω_C, and Ch_C is a probability measure on $\mathcal{F}_C$. For instance, if C picks out a certain type of die toss, Ω_C might be the set of possible outcomes $\{\mathit{One}, \mathit{Two}, \ldots, \mathit{Six}\}$, and $\mathcal{F}_C$ the set of subsets of Ω_C. Intuitively, the set $\{\mathit{Two}, \mathit{Four}, \mathit{Six}\}$ here stands for the unspecific property of landing with an even-numbered side up. I will often refer to the members of Γ as *conditions* and to the members of Ω_C as *(basic) outcomes*, but these names aren't meant to carry any significance: a condition may, for example, simply be a time.

The probability $Ch_C(A)$ of A relative to C is not derived from an unrelativised probability measure on a more inclusive algebra, perhaps as the ratio $Ch(A \wedge C)/Ch(C)$. We often have a well defined probability $Ch_C(A)$ e.g. of future states given present states, but no converse probability $Ch_A(C)$ of present states given future states, nor an unrelativised probability of the present state C. Relative to each index $C \in \Gamma$, there are ordinary conditional probabilities. Thus in the die toss example, it might be that $Ch_C(\mathit{Two}/\mathit{Even}) = Ch_C(\mathit{Two} \wedge \mathit{Even})/Ch_C(\mathit{Even}) = 1/3$. This must not be confused with $Ch_{C \wedge \mathit{Even}}(\mathit{Two})$ or $Ch(\mathit{Two}/C \wedge \mathit{Even})$, both of which are undefined. Confusing the two kinds of conditionality leads to what is known as "Humphreys' Paradox".

Traditional, single-case propositions are properties of a special kind: the proposition A is the property of being such that A. So the present framework does not rule out irreducible single-case chance. For the sake of generality, we might also allow for unrelativised chance. In this case, Γ is best identified with $\{\top\}$, the singleton of the tautologous property $\top$. A has unrelativised chance x iff A has chance x under condition $\top$. The condition is logically guaranteed to always obtain, so the relativisation does no work.

Now return to Coordination, Resiliency and the Principal Principle. We somehow

need to include the extra argument place for conditions. The most obvious way to adjust Coordination would be to replace $P(A/Ch(A)=x) \approx x$ with

(1) $\quad P(A/Ch_C(A)=x \wedge C) \approx x.$

This resembles the "reformulated" Principal Principle in [Lewis 1980: 97]. Lewis's principle says, in effect, that

(2) $\quad P(A/Ch=f \wedge C) = f_C(A),$

where $Ch=f$ is the proposition that the function f is the chance function, and the condition C is a complete history of a universe up to a certain time. [Meacham 2005] recommends generalising this Principle to other conditions, in order to accommodate chances in statistical mechanics and time-symmetrical versions of quantum mechanics. Like Lewis, Meacham assumes that the conditions Γ are very rich in information. (Meacham even assumes that each $C \in \Gamma$ uniquely determines the true chance function Ch.) (1) and (2) then guarantee a great deal of resiliency. To illustrate, suppose Γ is a finite set of hypotheses about the complete history of the world up to now. Let $Ch_@(A)=x$ be the proposition that the chance of A relative to the history of the (actual) world up to now is x. By the law of total probability, and the fact that $Ch_C(A)=x \wedge C$ is equivalent to $Ch_@(A)=x \wedge C$,

$$P(A/Ch_@(A)=x) = \sum_{C \in \Gamma} P(C)P(A/Ch_C(A)=x \wedge C).$$

By (1), $P(A/Ch_C(A)=x \wedge C) \approx x$ for all C. Hence if E is any disjunction of propositions in Γ – in other words, any information about the past – then $P(A/Ch_@(A)=x \wedge E) \approx x$. So information about chance screens off all information about the past.

It would be nice, I suppose, if we could always identify a chance function's domain of resiliency with the relevant conditions Γ. However, I do not want to assume that information about chance always screens off information about the relevant condition. I also don't want to assume that the conditions in Γ are informationally rich. If a statistical theory specifies probabilities for outcomes of coin tosses, then the relevant condition is *being a coin toss*, or perhaps *being a coin toss of type so-and-so*. The instantiation of this condition entails very little about the world, so Coordination does not automatically entail Resiliency.

In any case, there is something odd about (1) and (2). If A is *landing heads* and C *being a coin toss*, then (1) constrains a rational agent's prior degree of belief in *landing heads*, conditional on *being a coin toss*. But what does it mean to assign degrees of belief to a property?

[Lewis 1979] argued that the objects of credence really are properties. On this view, to assign high credence to a property is (roughly speaking) to self-attribute that property.

But this is not what we want. Most of us are fairly certain that we are not coin tosses, so it is not very useful to learn what we should believe conditional on this outlandish assumption. In general, our principles should cover attributions of the properties A and C to things other than ourselves: the probability that *this toss* will result in heads, given that *it* is a toss of the relevant type, should be such-and-such.

But now it matters how the relevant things are picked out. Suppose a certain coin will actually lands heads on its next toss. Then we can identify that toss as *the next toss of the coin resulting in heads*. The rational degree of belief in the hypothesis that the so-described toss will result in heads is 1, and remains 1 conditional on any hypothesis about chance. So Coordination fails. The problem here is that inadmissible information about the outcome has been smuggled into the way the chance process is picked out.

I will write '$R : A$' for the proposition that attributes the property A to the individual (event, process, etc.) identified by R. You may think of the *identifier* R as an individual concept or a Russellian definite description: '$R : A$' is true iff there is a unique individual that satisfies the condition R and this individual is A. I prefer to think of identifiers as binary relations, assuming with Lewis that the objects of credence are properties. When we attribute a property to an individual other than ourselves, we generally identify the individual by a relation it bears to us and our present location in space and time, as when we consider the *next* toss of *this* coin (see [Lewis 1979: sec.8]). For any (binary) relation R and property A, $R{:}A$ is the property that applies to an object x iff there is a unique object R-related to x and this object has A. For apparently non-relational ways of picking out an individual as, say, 'the tallest man in the history of the universe', the identifier is the relation that holds between x and y iff y is the tallest man in the history of the universe of x. Even more degenerate cases are "singular" identifiers $=_\alpha$ which always pick out a particular individual α. $=_\alpha$ is the relation that holds between x and y iff y is α.

Now we can take the Principal Principle to say that for many ordinary identifiers R and propositions B,

$$P(R{:}A \mid R{:}C \wedge Ch_C(A){=}x \wedge B) \approx x.$$

An identifier that picks out the next toss as 'next toss landing heads' would not count as suitable, because it contains inadmissible information.

The problem of inadmissible identifiers is rarely noted ([Skyrms 1980: 6ff.] is an exception), but it is a real phenomenon. For example, consider a variation of the Sleeping Beauty problem in which a second coin is tossed on Monday night. The proposition that *last night's coin* landed heads must then have different probability than the proposition that *the Sunday night coin* landed heads, even if last night's coin is the Sunday night coin (see [Schwarz 2012]). Inadmissible identifiers also often show up in "observation selection effects". Suppose the chance of life to evolve on an Earth-like planet is 0.1.

Conditional on this assumption, what is your subjective probability that life evolved on *the Earth-like planet on which you were born*? Not 0.1, of course. In this context, it is sometimes suggested that Coordination should be restricted to singular identifiers: if 'α' is a name of the Earth that conveys no qualitative information about its referent and its relation to ourselves, then conditional on the chance of life to evolve on an Earth-like planet being 0.1, the probability that life evolved on Earth-like planet α is plausibly 0.1 (see e.g. [White 2000]). However, it is controversial whether it is possible to pick out individuals in a relevantly non-descriptive way. Moreover, even the singular information $=_\alpha{:}C$ can reveal inadmissible information, because it indicates that there are many Cs: in worlds where almost everything is C, it is less surprising that the individual α is C than in worlds where almost nothing is C; but the information that there are many Cs can be evidence about the outcome of a chance process under condition C.

To keep issues of admissibility out of the Coordination condition, I will resort to a technical trick. I will explain how to transform any credence function P into a function P_C relative to which a certain identifier ϵC picks out an individual of type C without conveying any further information about the individual or the world. Intuitively, $\epsilon C\!:\!A$ is the proposition that an arbitrarily chosen C is A, and P_C is an extension of P that believes in a corresponding process of arbitrary choice. More formally, let W be the class of atomic possibilities ("worlds") in the domain of the credence function P. Relative to each $w \in W$, any property F has an extension F_w, i.e. the class of F instances that exist relative to w. For any condition C, let W_C be the class of all pairs $\langle w, c\rangle$ where $w \in W$ and $c \in C_w$. It may help to think of W_C as a class of *centred worlds*, except that the centre is not supposed to represent the location of an epistemic subject, but the "randomly selected" C: the 'you are here' arrow says 'the randomly selected C is here'. Note that W_C also excludes all worlds where there are no Cs. Since ordinary, uncentred propositions do not distinguish which C is selected, an ordinary proposition X is represented in W_C by the set of centred worlds $\langle w, c\rangle$ such that X is true at w. On the other hand, the centred proposition $\epsilon C\!:\!F$ is the set of centred worlds $\langle w, c\rangle$ such that c is in the extension of F at w.

If W_C is finite, we can now define P_C as the probability measure over W_C such that

$$P_C(\langle w, c\rangle) = \frac{P(w/\exists x C x)}{|C|_w},$$

where $|C|_w$ is the number of Cs at w. Thus P_C conditionalises P on the assumption that there are Cs and then evenly divides the probability of any world $w \in W$ among all the pairs $\langle w, c\rangle$. This ensures that every C-instance in a world has equal probability of being "selected". It follows that for uncentred propositions X,

$$(3) \qquad P_C(X) = P(X/\exists x C x),$$

and that $P_C(\epsilon C{:}A)$ equals the P-expectation of the ratio of As among Cs:

$$(4) \qquad P_C(\epsilon C{:}A) = \mathbb{E}_P\left[\frac{|A \wedge C|}{|C|}\right]$$

These equalities can also be used to define P_C (to the extent that we need it) for cases where W_C is infinite, as long as zero probability is given to the hypothesis that there are infinitely many instances of C. I will return to this limitation at the end of the paper.

Now Coordination can be expressed as follows.

> **Coordination.**
> Normally, if P is a rational prior credence function, A and C are properties, and $P_C(Ch_C(A){=}x) > 0$, then $P_C(\epsilon C{:}A/Ch_C(A){=}x) \approx x$.

Informally: the prior probability that an arbitrarily chosen C is A, given that the chance of A under C is x, should be approximately equal to x. Unlike in (1) and (2) above, there is no extra assumption about the instantiation of the condition C, since the randomly chosen individual of type C is already guaranteed to be an instance of C.

Coordination only indirectly links chance Ch_C to rational credence P by directly linking Ch_C to the C-transform P_C of P. Resiliency is also expressed in terms of P_C:

> **Resiliency.**
> Normally, if P is a rational prior credence function, A and C are properties, and $P_C(Ch_C(A){=}x) > 0$, then there is a substantial range of propositions B and identifiers R such that $P_C(R{:}A/R{:}C \wedge Ch_C(A){=}x \wedge B) = P_C(\epsilon C{:}A/Ch_C(A){=}x)$.

$P_C(X)$ equals $P(X/\exists x Cx)$ if X does not involve the identifier ϵC. Thus the Principal Principle, combining Coordination and Resiliency, can be expressed directly in terms of P, as promised above:

> **Principal Principle.**
> Normally, if P is a rational prior credence function, A and C are properties, and $P(Ch_C(A){=}x) > 0$, then there is a substantial range of identifiers R and propositions B such that $P(R{:}A/R{:}C \wedge Ch_C(A){=}x \wedge B) \approx x$.

The present formulations reduce to those of section 1 if C is the tautologous property $\top$ and R is any non-defective identifier, since $P_C(R{:}A)$ is then equivalent to $P(A)$.

4 Russell's Argument

Now return to the Russellian argument from section 2. On the frequentist interpretation, $Ch_C(A) = x$ says that the relative frequency of As within a suitable sequence of Cs is x.

Presumably this implies that the total number of Cs is positive and finite. Let P be a rational prior credence function for which $P(Ch_C(A) = x) > 0$. Let P' be P conditioned on $Ch_C(A) = x$. By (4), $P'_C(\epsilon C\!:\!A)$ is the expectation, by the lights of P', of the relative frequency of As among Cs. Since P' is certain that this ratio is x, $P'_C(\epsilon C\!:\!A) = x$. Hence $P_C(\epsilon C\!:\!A/Ch_C(A){=}x) = x$. We've proved Coordination.

Thus far, all we needed was the assumption that rational prior credence obeys the probability calculus. However, Coordination is only half of the story. We also need Resiliency. We have to show that there is (normally) a substantial range of ordinary propositions B and identifiers R such that $P_C(R : A/R : C \wedge B) \not\approx P(R : A)$, but $P_C(R{:}A/R{:}C{\wedge}Ch_C(A){=}x{\wedge}B) = P_C(\epsilon C{:}A/Ch_C(A){=}x)$. Together with Coordination, it then follows that $P(R{:}A/R{:}C \wedge Ch_C(A){=}x \wedge B) = x$. Showing this requires more substantial assumptions.

Take a concrete example. Suppose the relevant C-instance is picked out demonstratively, say, as 'the next toss of this coin'. If all you know is that the total ratio of heads among all tosses of a coin is 80%, what degree of belief should you assign to the hypothesis that the next toss will land heads? The probability calculus doesn't settle the answer. You might be certain that the next toss lands tails, or heads, or give equal credence to heads and tails. But recall that we are talking about *prior* credence. If any of these attitudes are part of your priors, they are either based on no evidence at all, or on the information that the relative frequency is 80%. In this case, wouldn't the attitude be irrational? If all you know is that the relative frequency of heads among some tosses is 80%, then you should be 80% certain that the next toss lands heads.

What's at work here is a principle of indifference. In general, consider any random-looking sequence of heads and tails with 80% heads. If the length of the sequence is n, then there are n possibilities about the location of the "next toss": it might be the first, or the second, ..., or the nth. In the absence of relevant evidence, you should give equal credence to these n possibilities. It then follows that your credence in the next toss landing heads will be 0.8. The principle of indifference required here is closely related to the principle of induction. To be confident that the tosses you are going to observe land tails while the unobserved tosses mostly land heads would reflect an irrational, counterinductive attitude towards the world.

It is notoriously difficult to find a satisfactory, precise formulation of indifference, or of inductive probabilities more generally. But these difficulties should not cast doubt on the fact that there are *some* indifference constraints on prior credence. Fortunately, the present argument requires only a very restricted, compartmentalised principle of indifference. It is not required that you distribute your credence uniformly over all ways things could be, which would presuppose a privileged parameterisation of logical space, and would arguably make it impossible to learn from experience. Imagine a grid imposed on logical space, each cell corresponding to a particular sequence of heads and tails. Set

aside all cells in which the ratio of heads is not 80%, as well as possibilities in which the distribution of heads and tails shows a conspicuous pattern. Each of the remaining cells divides into subcells, corresponding to different possibilities about the location of the next toss. We require that within each of the left-over cells from the original partition, you assign the same credence to every subcell. Nothing is said about how your credence should be divided between the larger cells, nor how it should be distributed within the subcells.

These remarks about *the next toss* carry over to other common identifiers such as *the previous toss* or *the toss presently reported by Jones*. They do not carry over to *the next toss that will land heads*. Here every specific hypothesis about a sequence of coin tosses still divides into different hypotheses about which of the tosses is the next toss that will land heads: the first, the second, and so on. But unless the sequence contains only heads outcomes, some of these subcells will be empty: if the ith element in the sequence is tails, then the ith element certainly isn't the next toss that lands heads. So your credence cannot be divided evenly between the subcells.

What about the extra information B in the Resiliency condition? For the neutral identifier ϵC, it is easy to see that ordinary, uncentred propositions are *always* admissible; i.e. for all uncentred propositions B, $P_C(\epsilon C : A \wedge Ch_C(A) = x \wedge B) = x$. On the other hand, ordinary identifiers and ordinary propositions together can become inadmissible. If B suggests that Jones tends to report only tails outcomes, then your credence in the outcome reported by Jones being heads, conditional on the relative frequency being 80% and B, won't be 0.8. We could say that in the presence of B, the identifier *the toss reported by Jones* is inadmissible. Or we could say that in the presence of this identifier, the information B is inadmissible. It doesn't really matter. (Technically, the addition of B in the Resiliency condition and the Principal Principle is redundant, since the information in B can always be folded into the identifier R.)

In section 2, we saw that information B about previous outcomes, combined with information about the total length of a sequence, can be inadmissible for frequentist chance. On the other hand, we also saw that if the credence function is not unduly opinionated, then Resiliency holds with respect to all B that specify not too many other outcomes. The relevant constraint on rational credence – exchangeability among random sequences – is another highly restricted form of indifference. As Lewis [1994: 229] points out, frequency information also tends to screen off many other facts that would otherwise be relevant to the outcome of the next toss, such as symmetries or asymmetries in the coin and the tossing procedure.

5 Best-System Probabilities

Let us move on to the leading Humean theory of chance: the best-systems approach. Here, 'chance' is defined indirectly via statistical theories. Let a *theory* be any logically closed set of sentences in a suitable language that includes resources to talk about probability. Given the total history H of (relevant) events in a world, theories can be ranked by their simplicity, strength, fit and possibly further criteria. Then chance is defined as the probability function employed in whatever theory ranks highest, on balance, in terms of these virtues. (See [Lewis 1994], [Loewer 2004], [Hoefer 2007] for more detailed expositions, and different ways of filling in the details.)

The *fit* between a theory T and a history H measures the extent to which T assigns high probability to events in H. Lewis suggested that if P^T is the probability function specified by theory T, then $P^T(H)$ can serve as measure of fit. This presupposes that statistical theories assign an absolute, unrelativised probability to complete histories. A natural generalisation to the present framework would use the product of $P^T_C(A)$ for each occurrence of an outcome A under a condition C in the history:

$$\prod_{\langle C,A\rangle \in H} P^T_C(A). \tag{5}$$

Here I assume that a history is represented as a sequence (or multiset) of condition-outcome pairs.

Formally, (5) defines a family of probability measures for every set of histories all of which agree in their frequency distribution over conditions $C \in \Gamma$. To make this more explicit, let T be any theory and P^T the probability function employed in T. Partition the space of histories by the distribution of frequencies over P^T's conditions Γ: $H \sim H'$ iff $|C|_H = |C|_{H'}$ for all $C \in \Gamma$. For any cell F in this partition, define $Fit_{T,F}$ as the probability measure over F given by

$$Fit_{T,F}(H) = \prod_{\langle C,A\rangle \in H} P^T_C(A).$$

$Fit_{T,F}$ resembles the "extended" frequentist chance function of section 2, but here it is not meant to represent T's probability for a history. Rather $Fit_{T,F}(H)$ is supposed to capture the extent to which H fits the probabilistic predictions of T, given the condition frequencies F.

What actually matters for the fit of a history to a theory are only the frequencies of outcomes in the history, not their order:

$$Fit_{T,F}(H) = \prod_{C\in\Gamma} \prod_{A\in\Omega_C} P^T_C(A)^{|C\wedge A|_H}.$$

Thus we can also measure fit directly in terms of a history's frequency distribution. Let $\langle k_1, \ldots, k_n\rangle_C$ be the set of histories $H \in F$ for which the outcomes $A_1, \ldots, A_n$ under

condition C have frequency $k_1, \ldots, k_n$, respectively. $Fit_{T,F}(\langle k_1, \ldots, k_n \rangle_C)$ is given by the multinomial formula

$$Fit_{T,F}(\langle k_1, \ldots, k_n \rangle_C) = \binom{|C|_F}{k_1, \ldots, k_n} \prod_{i=1}^{n} P_C^T(A_i)^{k_i},$$

where $|C|_F$ is the (constant) number of C instances in F histories. The fit for a single F history is the product of these values, for every condition C.

An alternative way to measure fit is to look at the differences $\Delta = |C \wedge A|_H - P_C^T(A)|C|_H$ between the "observed" frequencies $|C \wedge A|_H$ of A outcomes under condition C in a history and the "expected" frequencies by the light of the theory, $P_C^T(A)|C|_H$. Intuitively, the more the observed frequencies match the expected frequencies, the better the fit between theory and history. Aggregating the (squared normalised) differences Δ for all outcomes A under all conditions C yields

$$X^2 = \sum_{C \in \Gamma} \sum_{A \in \Omega_C} \frac{(|C \wedge A|_H - P_C^T(A)|C|_H)^2}{P_C^T(A)|C|_H}.$$

The lower X^2, the better the fit. On reflection, this measure is only plausible if the relevant frequencies are reasonably large. In this case, the X^2 value of a history H can be converted into an approximation of $Fit_{T,F}(H)$, so the two measures of fit are not really alternatives: since $Fit_{T,F}(\Delta = x)$ then follows an approximately normal distribution, $Fit_{T,F}(X^2 = x)$ approaches a sum of squared standard normal distributions; the χ^2 function with $\sum_{C \in \Gamma}(|\Omega_C| - 1)$ degrees of freedom thus yields an approximation of the $Fit_{T,F}$ of the set of histories in which the frequencies are at least as far from the expectation as in H. (The reasoning here parallels the reasoning behind the χ^2 test for "goodness of fit" in frequentist statistics.)

Now let P' be a rational prior credence function P conditioned on the assumption that (i) the frequency of Cs in actual history is k, and (ii) the best theory T assigns probability x to outcome A under condition C. Note that whatever the total frequency distribution F and best theory T might be, the distribution of $Fit_{T,F}$ over values of $|C \wedge A|$ is binomial with mean xk. Arguably, the P'-expectation of $|C \wedge A|$ should roughly equal this mean xk:

$$(6) \qquad \mathbb{E}_P(|C \wedge A| \;/\; |C| = k \wedge Ch_C(A) = x) \approx xk.$$

It follows that the P'-expectation of the relative frequency $\frac{|C \wedge A|}{|C|}$ roughly equals x. By definition, this expectation equals $P'(\epsilon C : A)$; so

$$(7) \qquad P(\epsilon C : A \;/\; |C| = k \wedge Ch_C(A) = x) \approx x.$$

If (7) is true for all k, we get the Coordination equality

$$(8) \qquad P(\epsilon C : A \;/\; Ch_C(A) = x) \approx x.$$

The only substantial assumption here is (6). Where does that come from? Recall that $Ch_C(A)=x$ says that the best theory assigns probability x to A under C. This reveals various facts about the pattern of events systematised by the theory. First of all, the pattern must be disorderly: otherwise the best theory would not involve probabilities. Moreover, the frequencies in the pattern must be distributed in such a way that some probability measure that assigns x to A under C strikes the best compromise, among measures on the relevant domain, between simplicity on the one hand and, on the other hand, closeness between probabilities and relative frequencies.

Now let T be an arbitrary theory that assigns x to A under C, and consider the histories (sequences of condition-outcome pairs) with k instances of C that are best systematised by T. In some of these histories, the number of $A \wedge C$ events might be much higher or lower than the theory's "expectation" xk; but then this lack of fit will usually be compensated by a closer match between frequency and expectation for other conditions and outcomes. By contrast, if $|A \wedge C|$ is fairly close to xk, then there is more leeway for the other frequencies to come apart from the expectation and still achieve high overall fit. As a consequence, there is in general a greater range of fitting histories in which $|A \wedge C|$ is fairly close to xk than histories in which the two are far apart. Moreover, the histories in which $|A \wedge C|$ is somewhat below xk will usually be matched by other histories in which $|A \wedge C|$ is somewhat above x. (6) assumes that the rational prior credence in either sort of deviation balances out so that the subjective expectation of the deviation is about zero.

Picture the binomial curve for $Fit_{T,F}$ over possible frequencies of A. If all you know is that T strikes the best balance between simplicity and maximising $Fit_{T,F}(H)$, then where do you think the frequency of A lies under the goodness of fit curve? In normal cases, you should believe that it is not too far from the maximum of the curve, and your credence in deviations on either side should balance out.

The rationality constraint reflected in (6) is again a restricted principle of indifference. The precise nature of the constraint depends on C and A, and on the criteria for a good theory. Not every instance is plausible. Suppose the best system assigns probabilities to informationally very rich events, such as the hypothesis A that the universe contains precisely 1000 coin tosses all of which land heads. The information that the best system assigns low probability x to A may then imply that A is false, since the best system of a world where A is true would not treat the coin tosses as chancy at all. So the P'-expectation of the frequency of A is zero, rather than xk. This is the phenomenon of undermining. As [Lewis 1980: 111f.] points out, Humean accounts of chance that allow for undermining propositions like A are incompatible with the strict Coordination condition. Accordingly, the present derivation of (8) breaks down for these propositions.

What about Resiliency? Here most of what I said for frequentism carries over. In fact, the best-systems account generally yields a wider domain of resiliency. That's because

best-system probabilities have to fit many frequencies, for many conditions. For example, if the state transitions for a certain system are modeled as a random walk, then the relative frequency of transitions from state C to state A may differ widely from the best system's probability $P_C(A)$ – especially if C or A is rare. Hence information about previous transition frequencies has little effect on how likely you should deem a transition from C to A, once you know the chance.

I have assumed that histories and outcome spaces are finite. If we lift this assumption, we run into the "zero-fit" problem (see [Elga 2004]). There are really two problems here, one arising from infinite outcome spaces, and one from infinite histories. Infinite outcome spaces are common in science, because outcomes are often real-valued. There are several ways to accommodate this in goodness of fit measures. A common method in statistics is to replace individual outcomes by reasonably chosen intervals, for example by partitioning the possible outcomes into $\sqrt{|C|}$ many intervals with uniform expected frequency.

Infinite histories are harder to deal with. If there are infinitely many instances of condition C in a history, our goodness of fit measure will no longer distinguish better from worse theories, since they all have zero fit. We also run into problems with the definition of P_C. A simple way around these issues might be to focus on finite subsets of Cs. If the world contains infinitely many Cs, we can look at increasingly large "samples", choosing all Cs within a certain distance from ourselves. If the world is well-behaved, the relative frequencies in these samples, and thereby the order of theories by fit, should converge. Of course, there is no logical guarantee that the world is well-behaved, but ill-behaved worlds deserve little rational credence, especially conditional on the hypothesis that the best system specifies probabilities relative to C.

6 Wrapping up

In order to say what chance is, we may first ask what chance does, and then see if we can find something that does that (compare [Lewis 1970: 22]). One thing chance does is constrain rational belief, in roughly the way expressed by Lewis's Principal Principle. We have seen that several Humean quantities fit this job description, including the "omniscient" probability function, relative frequencies in suitable sequences, and chance as characterized by best-systems accounts, no matter how various details in these accounts are spelled out.

One route to the conclusion that frequencies guide rational credence was reviewed in section 2. Here we saw, among other things, that if H_i says that the ith element in a (moderately long) sequence is of type H, and $Ch(H_i) = x$ is the proposition that the relative frequency of H in the sequence is x, then $P(H_i / Ch(H_i) = x) \approx x$ as long as P treats the sequence as exchangeable. In fact, it was enough that P assigns equal

probability to any two series of possible outcomes that agree in the total frequencies for each outcome type as well as their degree of orderliness. Under this condition, we could also verify that frequency information screens off other relevant information. In particular, $P(H_i/Ch(H_i) = x \wedge H_j) \approx x$. In section 4, I explored an alternative line of thought which draws on the idea that when we observe a member of a class X and know the proportion of Hs in X, then in the absence of further information our degree of belief that the observed X is H should equal that known proportion.

In section 5, I explained how these considerations carry over to best-systems accounts, where chances can to some extent diverge from relative frequencies so as to allow for a more compact statement of the relevant features of the world. Here the main idea was that our a priori credence in possibilities where the best-system probabilities lie below the frequencies should balance out with possibilities where they lie above, so that the *expected* relative frequency, given a certain hypothesis about the best-system probability, equals that probability.

All these arguments relied on non-trivial assumptions about rational credence: that in the absence of further information, agents should normally not deem a given sequence of outcomes more likely than an equally disorderly permutation; that on the information that most Xs are H, they should normally not be confident without any evidence that a given observed X is not-H; that on the information that the actual distribution of frequencies over a range of properties is best systematised by a function f, they should normally not be confident without further evidence that the true frequency for a given property lies below the value of f. These assumptions strike me as very plausible. Moreover, their plausibility does not rest on Humeanism or the interpretation of chance. Without assumptions like these, science could not be understood as a rational enterprise.

References

Robert Black [1998]: "Chance, Credence and the Principal Principle". *British Journal for the Philosophy of Science*, 49: 371–385

Jonathan Cohen and Craig Callender [2009]: "A Better Best System Account of Lawhood". *Philosophical Studies*, 145(1): 1–34

Bruno de Finetti [1937]: "La Prévision: ses lois logiques, ses sources subjectives". *Annales de l'Institute Henri Poincaré*, 7: 1–68

David Deutsch [1999]: "Quantum Theory of Probability and Decisions". *Proceedings of the Royal Society of London*, A455: 3129–3137

Persi Diaconis and David Freedman [1980]: "De Finetti's Generalizations of Exchangeability". In R. Jeffrey (Ed.) *Studies in Inductive Logic and Probability,* vol 2, chapter 11. Berkeley: University of California Press

Antony Eagle [Forthcoming]: "Probability and Randomness". In Alan Hájek and Christopher Hitchcock (Eds.) *The Oxford Handbook of Probability and Philoosphy,* Oxford: Oxford University Press

Adam Elga [2004]: "Infinitesimal Chances and the Laws of Nature". In Frank Jackson and Graham Priest (Eds.) *Lewisian Themes: The Philosophy of David K. Lewis,* Oxford: Oxford University Press, 68–77

Roman Frigg and Carl Hoefer [2010]: "Determinism and Chance From a Humean Perspective". In Friedrich Stadler et al. (Ed.) *The Present Situation in the Philosophy of Science,* Springer, 351–72

Ned Hall [1994]: "Correcting the Guide to Objective Chance". *Mind*, 103: 505–517

— [2004]: "Two Mistakes about Credence and Chance". *Australasian Journal of Philosophy*, 82: 93–111

Carl Hoefer [2007]: "The Third Way on Objective Probability: A Skeptic's Guide to Objective Chance". *Mind*, 116: 549–596

Colin Howson and Peter Urbach [1993]: *Scientific Reasoning.* La Salle (Ill.): Open Court Press, 2nd edition

David Lewis [1970]: "General Semantics". *Synthese*, 22: 18–67. Reprinted in Lewis's *Philosophical Papers*, Vol. 1, 1983.

— [1979]: "Attitudes *De Dicto* and *De Se*". *The Philosophical Review*, 88: 513–543. Reprinted in Lewis's *Philosophical Papers*, Vol. 1, 1983.

— [1980]: "A Subjectivist's Guide to Objective Chance". In Richard Jeffrey (Ed.), *Studies in Inductive Logic and Probability* Vol. 2, University of California Press. Reprinted in Lewis's *Philosophical Papers*, Vol. 2, 1986.

— [1994]: "Humean Supervenience Debugged". *Mind*, 103: 473–490

Barry Loewer [2004]: "David Lewis's Humean Theory of Objective Chance". *Philosophy of Science*, 71: 1115–1125

Christopher J. G. Meacham [2005]: "Three Proposals Regarding a Theory of Chance". *Philosophical Perspectives*, 19(1): 281–307

David H. Mellor [1971]: *The Matter of Chance*. Cambridge: Cambridge University Press

Bertrand Russell [1948]: *Human Knowledge: Its Scope and Limits*. London: George Allen and Unwin

Wolfgang Schwarz [2012]: "Lost memories and useless coins: revisiting the absendminded driver". Manuscript

Brian Skyrms [1980]: *Causal Necessity. A Pragmatic Investigation of the Necessity of Laws*. New Haven: Yale University Press

Michael Strevens [1999]: "Objective Probability as a Guide to the World". *Philosophical Studies*, 95: 243–275

David Wallace [2012]: *The Emergent Multiverse: Quantum Theory according to the Everett Interpretation*. Oxford: Oxford University Press

Roger White [2000]: "Fine-Tuning and Multiple Universes". *Noûs*, 34(2): 260–8211