Infinitesimal chances and the laws of nature*

Adam Elga

Penultimate draft, June 2003
Revised version to appear in
Australasian Journal of Philosophy

Abstract

The 'best-system' analysis of lawhood [Lewis 1994] faces the 'zero-fit problem': that many systems of laws say that the chance of history going actually as it goes—the degree to which the theory 'fits' the actual course of history—is zero. Neither an appeal to infinitesimal probabilities nor a patch using standard measure theory avoids the difficulty. But there is a way to avoid it: to replace the notion of 'fit' with the notion of a world being *typical* with respect to a theory.

1 The zero-fit problem

Take a god's eye view. Before you is the spacetime manifold: the distribution of local qualities to point-sized things and the spatiotemporal re-

*The author gratefully acknowledges help from Mark Colyvan, James Walmsley, Carl Hoefer, Nick Bostrom, Oystein Linnebo, Agustín Rayo, Daniel Nolan, Greg Restall, and an audience at the 1997 meeting of the Australasian Philosophical Association. Special thanks go to Alan Hájek, who provided detailed feedback at several stages.

lations among those things. According to the thesis of Humean supervenience, everything that is true of the actual world is made true somehow by this arrangement [Lewis 1986b, ix].[1] In order to countenance laws of nature, a defender of the thesis is obliged to say how it is that this arrangement determines what laws there are.

One proposal, the *best-system analysis of lawhood* [Lewis 1994], draws our attention to various deductive systems. Each system makes only true assertions about what happens, and (perhaps) also makes assertions about the chances of various things happening in various circumstances.

On this proposal, the laws are the regularities that are members of the best candidate system, and the chances are whatever the best candidate system asserts them to be. The best system is the one with the best balance of the following three virtues. *Simplicity:* A system is simple to the extent that it can be concisely formulated in a certain canonical language.[2] *Strength:* The strength of a system is its informativeness, both regarding matters of particular fact and regarding what chances arise in various circumstances. *Fit:* Systems that assign chances to certain courses of history also assign a chance to the actual course of history. The *fit* of such a system

[1] I have omitted several qualifications to the thesis of Humean supervenience. The qualifications do not affect the present discussion. See for example [Lewis 1994, 474-475].

[2] The language is one with a primitive predicate for each perfectly natural property. See [Lewis 1983, 367-368].

is defined to be that chance—how likely the system counts it that things would go just as they actually do.[3] (By stipulation, systems that don't mention chances have perfect fit.)

Here is how the analysis works in a simple case. Consider a world that consists of a long finite sequence atomic events, each of which happens in one of two ways. Call the events 'tosses', and call the ways 'heads' and 'tails'. At one extreme, the entire sequence may be captured by a simple regularity (say, heads and tails alternate throughout). In that case, a system stating that regularity will be simple, strong, and fit well—it will be the best system. As a result, that regularity will qualify as a law.

At another extreme, the sequence may not contain many simply describable regularities. In that case, any system that is very informative regarding the details of the sequence will need to be extremely complicated. Certain systems avoid this complication by asserting that the sequence is the result of a repeated chance process. Such systems gain much in simplicity at the cost of some strength (they only address the chances of various sequences, as opposed to making claims about the details of the actual

[3]Some systems specify the chances of future evolutions *given an initial condition*, but fail to specify a chance distribution over initial conditions. The above definition leaves the fit of such systems undefined. A natural fix is to stipulate that the fit of such a system is the conditional chance that the system ascribes to the actual course of history, given that the universe started in the initial state that it did.

sequence). We can compare these chance-ascribing systems by comparing their fits.

For example, suppose that 1/10 of the tosses land heads, and that the heads outcomes are scattered haphazardly among the tosses. One candidate system asserts that the tosses are independent chancy events, and that each has chance 1/10 of landing heads. A competing system also treats the tosses as independent, but asserts that each toss has chance 1/2 of landing heads. The two systems are roughly equally informative and simple, but the 'chance-1/10' system ascribes a higher chance to the actual sequence than the 'chance-1/2' system does. In other words, the 'chance-1/10' system fits better. It is thereby a better competitor in the simplicity/strength/fit competition. Plausibly, this system beats all comers and thereby qualifies as being the correct system of laws.

That is a welcome result for the best system analysis. A friend of Humean supervenience is committed to thinking that the actual sequence of outcomes determine the chances. Given that commitment, it makes sense that a tails-heavy sequence of tosses makes it the case that each toss has a high chance of landing tails.

Notice that in this case, the notion of 'fit' deserves its name. The chances ascribed by the 'chance-1/10' theory accord well with a history in which

1/10 of the tosses land heads. The chances ascribed by the 'chance-1/2' theory do so poorly. And the notion of 'fit' measures that difference. More generally, when finite state-spaces are involved, the notion of fit usefully ranks competing candidate systems of chancy laws. But—as I learned from Ned Hall [1996], and as Nick Bostrom has independently noted [1999]—when infinite state-spaces are involved, the notion of fit no longer ranks chancy systems in a useful way.

To see this, change the above heads/tails example by letting the sequence of tosses be (countably) infinite. Much of the best-system analysis goes through unchanged. Again, if the entire sequence is captured by a simple regularity, then a system stating that regularity will be decisively best. Again, if the sequence is unpatterned, then any system that is very informative regarding the details of the sequence will need to be extremely complicated. And again, certain competing systems avoid this complication by asserting that the sequence is the result of a repeated chance process.

But when we compare these chancy systems by comparing their fits, we run into trouble. We might hope that the systems that fit best are the ones whose chances accord well with the actual pattern of outcomes. But our hopes are dashed: *far too many systems have fits equal to exactly zero*.

For example, suppose that in the infinite sequence of tosses, 1/10 land heads.[4] In order for the notion of fit to do its job, the 'chance-1/10' system should fit this sequence better than the 'chance-1/2' system does. (Recall that the 'chance-1/10' system treats the tosses as independent chancy events with chance 1/10 of heads on each toss.) But that's not so: *both* systems ascribe chance zero to the sequence, and so both systems have fits equal to zero.[5] More generally, when continuously infinite state-spaces are involved, a great many candidate systems (including systems of chancy laws that physicists have taken seriously) will ascribe zero chance to any individual history.[6] That leaves the best-system analysis with no way of differentiating between chance-ascribing systems whose chances accord well with the actual history, and those whose chances do so poorly. This is the *zero fit problem*.

[4] 1/10 of the tosses land heads in the sense that the limiting relative frequency of heads is 1/10.

[5] Proof: Consider the chance-1/2 system (the argument in the chance-1/10 case is similar). For any natural number n, consider the proposition E_n that specifies the actual outcomes of the first n tosses. The proposition E that specifies *all* of the outcomes is stronger than each E_n, and so its chance can't be any greater than the chance of any E_n. But the chance of E_n is 2^{-n}, which gets arbitrarily close to zero as n tends to infinity. So the chance of E—and hence the fit of the chance-1/2 system—equals zero.

[6] For example, any system that treats an infinite series of events as a series of (nontrivial) independent chancy coin-tosses will have a fit of zero. So will any theory of radioactive decay according to which the chance of decay within a time interval is gotten by integrating a density over that interval.

2 Infinitesimals to the rescue?

It can seem odd when a system counts an outcome as possible, and yet ascribes it chance zero.[7] There is a way of using nonstandard models of analysis to avoid this oddness. Nonstandard extensions of the real line contain *infinitesimals*—positive numbers smaller than any positive real number. And the nonstandard universe contains *nonstandard probability functions*, which take their values from a nonstandard extension of the reals.[8]

Nonstandard probability functions may ascribe infinitesimal probability to certain outcomes. As a result, if we allow candidate systems to be associated with nonstandard probability functions, we may impose the *regularity condition*: that each candidate system ascribe some nonzero chance to every outcome that it counts as possible. In doing so, we may hope to rescue the best system analysis from the zero fit problem. For in that case, no system under consideration will have a fit of zero.[9]

[7] With standard probability functions, this is often unavoidable, since such functions assign positive probability to at most countably many incompatible propositions.

[8] For an accessible introduction to nonstandard analysis, see [Skyrms 1980, Appendix 4]. For a brief technical introduction, see [Bernstein and Wattenberg 1969, Section 1]. For a thorough technical introduction, see [Hurd and Loeb 1985].

[9] Perhaps Lewis had this strategy in mind. He insists that 'Zero chance is *no* chance, and nothing with zero chance ever happens. The [fair spinner's] chance of stopping exactly where it did was not zero; it was infinitesimal, and infinitesimal chance is still *some* chance.' [Lewis 1986a, 176] and writes 'It might happen—there is some chance of it, infinitesimal but not zero—that each nucleus lasted for precisely its expected lifetime...' [Lewis 1986c, 125]. He also explicitly invokes nonstandard probability functions in another circumstance in which zero probabilities threaten to cause trouble [Lewis 1986c,

Let us apply this strategy to the special case mentioned above: a world consisting of an infinite sequence of tosses. If the strategy fails in this case—a simple instance in which the zero fit problem arises—it is sure to fail in general.

And it does fail in this case. To see why, consider the *Bernoulli systems*—candidate systems that treat the tosses as independent chancy events. For any real number strictly between zero and one, let B_x be the probability function that treats the tosses as independent events with probability x of heads. These functions are all ruled out as chance functions by the regularity requirement. (They assign probability zero to each individual infinite sequence of toss outcomes). What we need are regular nonstandard probability functions to play the role that the functions B_x play in the standard case.

We can be assured that there are such functions. It turns out that for every standard probability function there exists a nonstandard probability function that (i) approximates the standard probability function (in the sense that the two functions never differ by more than an infinitesimal) and (ii) satisfies the regularity condition (i.e., assigns positive probability to every nonempty proposition). Appendix A uses a method due to

88–90].

Vann McGee to prove the existence of such functions. Once we have imposed the regularity condition, such functions will represent the chances ascribed by candidate systems.[10]

So for each B_x there exists regular nonstandard probability function B'_x that approximates B_x. So far, so good. Now suppose that in fact 1/10 of the tosses land heads. Which functions fit this sequence of outcomes? If things turn out well, then the functions B'_x with good fits will be the ones for which x is close to 1/10. Unfortunately, things do not turn out well. The trouble is that for each B_x, there are *many* regular nonstandard probability functions that approximate it. Furthermore, the functions that treat the coin as a chance device with bias 1/10 do *not* fit any better than the ones that treat it as a chance device with any other bias. For example, it can happen that a nonstandard probability function $B'_{1/2}$ that approximates $B_{1/2}$ fits *much better* than a function $B'_{1/10}$ that approximates $B_{1/10}$—even though the actual limiting relative frequency of heads is 1/10.

[10]To my knowledge, there are only two methods for cooking up regular nonstandard probability functions. One method—mentioned in the text above and explained in detail in Appendix A—is to start with an arbitrary (standard) probability function and build a nonstandard regular probability function that approximates it. A second method is to focus on a probability space with some natural symmetries, and seek regular probability functions on that space that respect those symmetries. (For example, [Bernstein and Wattenberg 1969] proves the existence of a nonstandard probability function on the unit circle that (i) is rotationally invariant (up to an infinitesimal) and (ii) assigns the same (infinitesimal) probability to each point. [Parikh and Parnes 1972] extends Bernstein and Wattenberg's construction.) This second method treats only certain very special cases, and so fails to rescue the best-system analysis from the zero-fit problem.

The trouble is this. We have required our nonstandard probability function to be regular, and to approximate given standard probability functions. But those requirements only very weakly constrain the probabilities those functions assign to any individual outcome. (This follows from Corollary 1, which is proved in Appendix A.) And the fit of a system associated with such a function is just the chance it assigns to actual history. So the fit of such a system indicates nothing about how well its chances accord with actual history.[11]

The moral is that if we allow nonstandard probability functions, goodness of fit is no indication that a system's chances accord well with actual frequencies. In that case, comparing fits still fails to distinguish between systems whose chances accord with the actual frequencies and systems whose chances do so poorly. So introducing nonstandard probability functions is of no help to the best-system analysis.

[11] It might be thought that simplicity saves the day. Perhaps $B'_{1/10}$, though it fits worse, is much more simple than $B'_{1/2}$. But since the proof that assures us of the existence of both functions is nonconstructive, we have no reason to expect that to be so.

3 An alternate proposal: fit profiles

One might hope to use a standard measure-theoretic technique (the method of integrating over densities) to solve the zero fit problem. Sadly, such methods do not help. So I relegate their consideration to Appendix B, and turn to an alternate approach.

The fit between a system S and a world w is defined to be the chance the system ascribes to a certain proposition: the proposition true only at w. Separate out two parts of this definition:

1. Good fit between a system S and a world w is a matter of certain propositions true at w (call them 'test propositions') having high enough chance, according to S.

2. There is exactly one test proposition: the proposition true only at w.

The conjunction of these parts leads to the zero fit problem. So let us modify the notion of fit by abandoning part 2. We're left with the task of specifying the test propositions—the task of saying *which* propositions true at w are the ones whose chances (according to S) determine the fit between S and w.

Some work by Gaifman and Snir helps us out here [1980]. Gaifman and Snir formalize the notion of a world being typical with respect to a probability function.[12] Given a world w, a probability function P, and a set T of test propositions, they stipulate that w is typical with respect to P and T iff P assigns nonzero probability to every test proposition true at w. Their research makes it reasonable to let the test propositions be those propositions simply expressible in a certain first-order language.

For example, consider just worlds that are infinite heads/tails sequences, and fix a language L in which it is natural to describe those sequences.[13] Let the test propositions be the propositions simply expressible[14] in L and examine the worlds Gaifman and Snir's criterion counts as typical with respect to $B_{1/2}$. With this choice of test propositions, Gaifman and Snir's criterion does just the right thing. That is, the worlds that Gaifman and Snir's criterion counts as typical with respect $B_{1/2}$ satisfy just the sorts of conditions we'd expect a world to satisfy if it consisted of a sequence of independent unbiased chance events. (Examples of such conditions: that the limiting relative frequency of heads is $1/2$; that in the limit the

[12]Gaifman and Snir's work is much more general than my discussion might indicate. I have extracted a small piece for the sake of clarity. I have also modified their terminology slightly: where they write that a world is *random* with respect to a probability function, I write that it is *typical* with respect to that function.

[13]An appropriate language is first-order arithmetic augmented with a one place predicate H, where $H(i)$ is to be interpreted as 'Toss i is heads'.

[14]More exactly, let the test propositions be the ones expressible by Σ_2 sentences of L—sentences of the form $\exists x \forall y \phi$, where ϕ is replaced by a formula containing only bounded quantifiers [Gaifman and Snir 1980, 501].

pattern HTH appears exactly as often as THT; and that deleting the odd-numbered elements of the sequence produces another sequence that satisfies the above conditions.)

In general, by fixing the choice of test propositions as the simply expressible propositions, Gaifman and Snir's criterion succeeds in picking out those worlds that are typical relative to a probability function. Now, the *job* of the notion of fit in the best-system analysis is to measure the typicality of a world relative to a system. So let us adapt Gaifman and Snir's criterion and say that the test propositions figuring in our revised notion of fit are exactly the true, simple propositions. Accordingly, the way to compare chance-ascribing systems for fit is to compare their *fit profiles*: the chances they ascribe to the test propositions. System X fits better than system Y iff the chances X assigns to the test propositions are predominantly greater than the corresponding chances that Y assigns. (Perhaps X assigns a higher chance than Y to every test proposition. Or perhaps X assigns higher chances than Y overall—in a sense I won't try to make precise—although not to every single test proposition.) More generally, the fit of a system is an overall measure of the magnitude of the chances it ascribes to true, simple propositions.

This proposal avoids the zero fit problem. For even systems that assign chance zero to actual history will assign positive chance to many test propositions. So even in this case, the fit profile of a system remains a useful indicator of how well its chances accord with the actual outcomes.

On this proposal, it can certainly happen that two systems are incomparable with respect to fit. That is no special worry—the best-systems analysis already depends on the hope that some system will be robustly best, as regards the tradeoff between simplicity, strength, and fit. It is no great cost to add an additional hope: that this robustly best system possess a fit profile that holds its own against the profiles of its competitors on any reasonable way of judging when one profile assigns higher chances overall than another.

Appendices

A Nonstandard probability functions

This appendix proves the existence of a nonstandard probability function that (i) approximates any given conditional probability function (ii) assigns strictly positive probability to every nonempty proposition. The method is a variant of one due to Vann McGee [1994]. One might hope that such functions save the best-system analysis from the zero-fit problem. For if such functions represent the chances ascribed by competing systems, each system will have strictly positive fit.

Corollary 1 (proved below) dashes any such hopes. The trouble is that once infinitesimals are in play, the notion of fit fails to measure the degree to which a system's chances accord with the actual pattern of outcomes. For example, consider any standard probability function that ascribes probability zero to actual history. There are *many* regular nonstandard probability functions that approximate this function. All of them assign infinitesimal probability to actual history. But the probabilities that these approximating functions ascribe to actual history span the *entire* range of infinitesimals (that's what Corollary 1 says). So by picking an appropriate approximating function, we can get any such system to have any (infinitesimal) fit we'd like.

Definitions and notation. A probability space is a pair $\langle W, \mathcal{F} \rangle$, where W is a set of elementary outcomes and $\mathcal{F}$ is an algebra of subsets of W, called events. $\mathcal{F}'$ is the set of nonempty events. A (finitely additive) probability function is a function P that maps $\mathcal{F}$ to the unit interval in such a way that (i) $P(W) = 1$ and (ii) $P(A_1 \cup \cdots \cup A_n) = P(A_i) + \cdots + P(A_n)$ whenever $A_1, \cdots, A_n$ are disjoint events.

A nonstandard probability function is a function P that maps $\mathcal{F}$ to a nonstandard extension of the unit interval in such a way that those same conditions are satisfied. A probability function is *regular* iff it assigns positive probability to every nonempty event. If x and y are members of a nonstandard extension of the real line, '$x \cong y$' means that x and y differ by at most an infinitesimal. A *conditional probability function*[15] on the probability space $\langle W, \mathcal{F} \rangle$ is a function $C : \mathcal{F} \times \mathcal{F}_{\prime} \to [\prime, \infty]$ such that

1. For any $D \in \mathcal{F}$, $C(\cdot, D)$ is a probability function with $C(D, D) = 1$.

2. For any $X, Y, Z \in \mathcal{F}$ with $Y \cap Z$ nonempty, $C(X \cap Y, Z) = C(Y, Z) \cdot C(X, Y \cap Z)$.

A nonstandard probability function P' is said to *approximate* a probability function P iff for any $X \in \mathcal{F}$, $P'(X) \cong P(X)$. P' is said to approximate a conditional probability function C iff for any $X \in \mathcal{F}$ and $Y \in \mathcal{F}_{\prime}$, $P'(X|Y) \cong C(X, Y)$. The following theorem was proved as Theorem 3.4 of [Krauss 1968] and independently rediscovered as Theorem 1 of [McGee 1994].

Theorem 1 *For any conditional probability function C, there exists a regular nonstandard probability function that approximates C.*

It will be handy to have the following slightly stronger result.

Theorem 2 *For any conditional probability function C and infinitesimal α, there are regular nonstandard probability functions P^+ and P^- that satisfy the following conditions:*

1. *P^+ and P^- both approximate C.*

2. *For any nonempty $B \in \mathcal{F}$,*
   $P^+(B) > \alpha$.

3. *For any $B \in \mathcal{F}$ such that $C(B, W) = 0$,*
   $P^-(B) < \alpha$.

Proof A small modification of the proof in [McGee 1994] does the trick. Let relation Y^+ hold between $D \in \mathcal{F}_{\prime}$ and a nonstandard probability function P iff

(1) For all $B \in \mathcal{F}$, $P(B|D) \cong C(B, D)$.

(2$^+$) $P(D) > \alpha$.

[15] On conditional probability functions, see [Renyi 1955 Popper 1952 Hájek 2003].

Let relation Y^- hold between $D \in \mathcal{F}_{\prime}$ and a nonstandard probability function P iff

(1) For all $B \in \mathcal{F}$, $P(B|D) \cong C(B, D)$.

(2^-) For any $B \in \mathcal{F}$ such that $C(B, W) = 0$,
$P^-(B) < \alpha$.

Showing that Y^+ is finitely satisfiable is sufficient to show the existence of a nonstandard model containing a function P^+ meeting the conditions of the theorem (see, e.g., [Hurd and Loeb 1985]). Similarly, showing that Y^- is finitely satisfiable is sufficient to show the existence of a nonstandard model containing the required function P^-.

First we show that Y^+ is finitely satisfiable. Suppose that we are given **D**, a finite subset of $\mathcal{F}_{\prime}$. We need a nonstandard probability function P such that $Y^+(D, P)$ for every $D \in \mathbf{D}$. To that end, form the decreasing sequence of members of $\mathcal{F}$ as follows: At stage 0, set $A_0 = W$. At stage k, set $A_{k+1} = \bigcup \{D \in \mathbf{D} | C(D, A_k) = 0\}$. Let m be the greatest integer such that A_m is nonempty.

Choose an infinitesimal ϵ such that $\epsilon^{m+1} > \alpha$. Define P by stipulating that for any $B \in \mathcal{F}$,

$$P(B) \stackrel{\text{def}}{=} k^{-1} \left(C(B, A_0) + \epsilon C(B, A_1) + \epsilon^2 C(B, A_2) + \cdots + \epsilon^m C(B, A_m) \right),$$

where $k = \sum_{i=0}^{m} \epsilon^i$ is a normalizing constant. P is a convex combination of probability functions, and hence is a probability function. Now take any $D \in \mathbf{D}$ and $B \in \mathcal{F}$. Let z be the least integer such that $C(D, A_z) > 0$. Then: $P(B|D) = P(B \cap D)/P(D) = (\sum_{i=z}^{m} \epsilon^{i-z} C(B \cap D, A_i)) \,/\, (\sum_{i=z}^{m} \epsilon^{i-z} C(D, A_i))$ $= (C(B \cap D, A_z) + \text{ infinitesimal}) \,/\, (C(D, A_z) + \text{ infinitesimal}) \cong C(B, D)$. This shows that P satisfies the first condition of Y^+ with respect to any $D \in \mathbf{D}$. To see that P satisfies (2^+), start with any $D \in \mathbf{D}$, and let z be the least integer so that $C(D, A_z) > 0$. Then $P(D) = k^{-1} \sum_{i=0}^{m} \epsilon^i C(D, A_i)$ $= k^{-1} \sum_{i=z}^{m} \epsilon^i C(D, A_i) \geq \epsilon^z k^{-1} C(D, A_z) \geq \epsilon^{z+1} > \alpha$ This shows that Y^+ is finitely satisfiable.

To show that Y^- is finitely satisfiable, repeat the same construction, except that instead of choosing ϵ so that $\epsilon^{m+1} > \alpha$, choose ϵ so that $\epsilon < \alpha$.
□

Corollary 1 *For any probability function P on $\langle W, \mathcal{F} \rangle$ such that for all $w \in W$, $P(\{w\}) = 0$, and for any infinitesimal α, there are regular nonstandard probability functions P^+ and P^- that satisfy the following conditions:*

1. *P^+ and P^- both approximate P.*

2. *For any $w \in W$, $P^+(\{w\}) > \alpha$.*

3. *For any $w \in W$, $P^-(\{w\}) < \alpha$.*

Proof Let $\prec$ be a well-ordering of W. Define a conditional probability function C by stipulating that for any $X \in \mathcal{F}$ and $Y \in \mathcal{F}'$: $C(X, Y) = P(X \cap Y)/P(Y)$ if $P(Y) > 0$; $C(X, Y) = 1$ if $P(Y) = 0$ and X contains the $\prec$-minimal element of Y; and $C(X, Y) = 0$ otherwise.

Since $C(\cdot, W) = P(\cdot)$, in order to approximate P it is sufficient to approximate C. So apply Theorem 2 to C and α to obtain P^+ and P^- as desired. □

B Integrating over densities

One might hope to use the method of integrating over densities to solve the zero fit problem. This appendix presents the most natural such proposal, and shows how it comes to grief.

Here is the proposal. Instead of associating each candidate system with a probability function, associate each system with a probability *density* function—a function that assigns a density (a real number) to each world. Define the chance that a system ascribes to a proposition to be an appropriate average of the densities it assigns to worlds at which that proposition is true. (In other words, the probability function to be associated with a system is the one gotten by integrating over the system's density function in the usual way.) Finally, stipulate that the fit of a system is the density it assigns to the actual world. This setup seeks to avoid the zero fit problem by allowing a system to assign nonzero *density* to the actual world (and hence have nonzero fit), even if it assigns zero *chance* to the actual world.

Unfortunately, this setup doesn't work. Trouble arises because in order to recover a probability function from a probability density function, one needs an underlying probability measure with respect to which one performs an integration. And one had better use the *same* measure when recovering probability functions from the probability density functions associated with different systems—otherwise the relative density two systems assign to the same world would have no significance.

So we need to fix a common underlying measure over the space of possible worlds. But no appropriate choices are available. To see this, consider how the story would have to go in order for us to check how well the various Bernoulli systems fit a world consisting of an infinite sequence of tosses. In order to get started, we'd need an underlying measure μ on the space of infinite heads/tails sequences, as well as a probability density function f_x to associate with each Bernoulli system with bias x. We would need, for each f_x, that one could recover the probability function B_x by integrating over the density f_x in the usual way (otherwise the Bernoulli systems wouldn't deserve their names). But no choice of a measure μ and functions f_x satisfies these conditions.[16] As a result, even in this simple case, the integrate-over-densities proposal doesn't get off the ground.

References

Bernstein, A. and F. Wattenberg 1969. Nonstandard measure theory. In Luxemburg, W. A. J., ed., *Applications of Model Theory*, Holt, Rinehard and Winston.

[16]Proof: Assume for contradiction that some choice of underlying probability measure μ and density functions f_x *does* satisfy these conditions. That is, assume that for all $x \in (0,1)$, B_x can be recovered by integrating over the density f_x in the usual way. That is, for all $x \in (0,1)$, for all μ-measurable A, $B_x(A) = \int_A f_x d\mu$. For each $x \in (0,1)$, let L_x be the set of worlds whose limiting relative frequency of heads is x. The law of large numbers tells us that for each x, $B_x(L_x) = 1$. By assumption, $B_x(L_x) = \int_{L_x} f_x d\mu$, and so $\mu(L_x) > 0$. Note that L_x and L_y are disjoint whenever $x \neq y$. But now μ assigns nonzero probability to each of a continuum of pairwise disjoint propositions. That's impossible because a probability measure assigns nonzero probability to at most countably many pairwise disjoint propositions.

Bostrom, Nick 1999. A critique of David Lewis's theory of chance? Manuscript.

Gaifman, Haim and Marc Snir 1980. Probabilities over rich languages, testing and randomness. *Journal of Symbolic Logic* 47(3): 495–548.

Hájek, Alan 2003. On what conditional probability could not be. *Synthese* Forthcoming.

Hall, Ned 1996. Personal communication.

Hurd, Albert E. and Peter A. Loeb 1985. *An Introduction to Nonstandard Real Analysis*. New York: Academic Press.

Krauss, P. H. 1968. Representation of conditional probability measures on boolean algebras. *Acta Math. Acad. Scient. Hungaricae* 19: 229–241.

Lewis, David 1983. New work for a theory of universals. *Australasian Journal of Philosophy* 61(4): 343–377.

Lewis, David 1986a. Causation. In *Philosophical Papers*, Oxford: Oxford University Press, volume 2.

Lewis, David 1986b. *Philosophical Papers*, volume 2. New York: Oxford University Press.

Lewis, David 1986c. A Subjectivist's Guide to Objective Chance. In *Philosophical Papers*, Oxford: Oxford University Press, volume 2.

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

McGee, Vann 1994. Learning the Impossible. In Eells, Ellery and Brian Skyrms, eds., *Probability and Conditionals: belief revision and rational decision*, Cambridge: Cambridge University Press, chapter 10.

Parikh, Rohit and Milton Parnes 1972. Conditional probabilities and uniform sets. In Dold, A. and B. Eckmann, eds., *Probabilistic Analysis and Related Topics*, New York: Springer-Verlag, number 369 in Lecture Notes in Mathematics, pages 180–194.

Popper, Karl R. 1952. *The Logic of Scientific Discovery*. New York: Basic Books, second edition.

Renyi, A. 1955. On a new axiomatic theory of probability. *Acta Math. Acad. Scient. Hungaricae* 6: 285–335.

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