Inferring Probabilities from Symmetries
Author(s): Michael Strevens
Source: *Noûs*, Vol. 32, No. 2 (Jun., 1998), pp. 231-246
Published by: Blackwell Publishing
Stable URL: http://www.jstor.org/stable/2671966
Accessed: 17/12/2009 22:40

NOÛS 32:2 (1998) 231–246

Inferring Probabilities From Symmetries

MICHAEL STREVENS
Stanford University

Consider a perfectly symmetrical dodecahedron (see Figure 1). Paint one face red, and it may be used as a die. What is the probability that such a die, when tossed, lands red face up? Without having touched, or even laid eyes on, such an object, it is possible to give a firm answer: the probability is one in twelve. Find such a die, and experience will bear out this assertion. The red face will land uppermost roughly one twelfth of the time. We have made, apparently, an a priori inference to a fact about the world.

How is it that we are able to infer the probability with such success? The conventional answer to this question is that we invoke a rule called the Principle of Indifference (or the Principle of Insufficient Reason).[1] This rule tells us that, in the absence of any known reason to assign two events differing probabilities, they ought to be assigned the same probability. Applied to the dodecahedral die, the principle allows us to reason as follows. The die may land with any of twelve faces uppermost. We know nothing that distinguishes these twelve possibilities, so we must assign each an equal probability. Since the sum of these probabilities must equal one (some face must be uppermost), each possibility should be assigned a probability of one twelfth.

This explanation of our probabilistic prowess cannot be right. The Principle of Indifference is not even the right *kind* of rule to explain our successful inference. It purports to tell us which probabilities are rational given our ignorance, helping us to do the right thing under adverse epistemic circumstances. But our achievement is not mere rationality; it is truth. We infer the *correct physical* probability for the event in question. This suggests that our inference is not based on ignorance at all, and thus not based on a principle of "indifference" or "insufficient reason". Two arguments support this suggestion.

First, it is surely the case that we can never reliably get from ignorance to truth, because the nature of the world is independent of our epistemic deficits. The fact we do not know anything about A does not constrain the way things are with A.

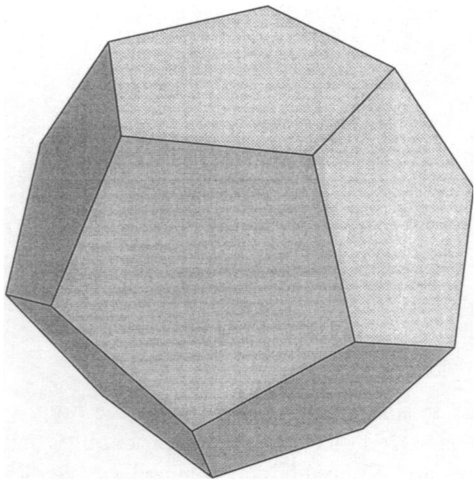

Figure 1. A dodecahedral die

Second, in the sort of case under discussion, we have an inferential disposition that is incompatible with the claim that our inference is based on ignorance. The disposition is as follows. Suppose that you are certain that some particular dodecahedral die is absolutely symmetrical, externally and internally (an internally symmetrical die is not "loaded"). Suppose also that, on repeated throwing, the red face of this die lands uppermost one half of the time. Should you revise your estimate of the probability? Not at all. Given the symmetry of the die, the probability of a red face must be one twelfth. That the frequency of red faces is one half can only be an accident. This inferential disposition—our refusal to revise the probability in the light of the new data—suggests that our knowledge of the probability, far from being based on ignorance, is based on knowledge so certain that it overrides any amount of information about outcomes. In the case described, that can only be our certain knowledge of symmetries.[2]

The knowledge from which we infer the probability, then, is our certain knowledge of the symmetry of the die. It is the purpose of this paper to explain, and to justify, the rule of inference that takes us from knowledge of physical symmetries to knowledge of actual physical probabilities. (I stress once more that this rule is very different from the Principle of Indifference, which takes us from symmetries in our *knowledge*—or more exactly, ignorance—to *rational* probabilities.[3]) I call the inference from symmetries to probabilities *non-enumerative statistical induction* (NEI).[4] *Enumerative statistical induction*, by contrast, is that form of reasoning in which we infer probabilities from frequencies of outcomes, as when we conclude that smoking causes cancer with a certain probability by perusing the relevant medical statistics.

The paper has four parts. In the first part I show that non-enumerative statistical induction plays an important role in several sciences. Thus the problem

extends beyond the gaming halls. The second part considers the Principle of Indifference in more detail, the third an alternate account of NEI due to Henri Poincaré and Hans Reichenbach. The final part presents my own account of NEI.

The Success of Non-Enumerative Statistical Induction

It might be thought that successful NEI is confined to a few unimportant cases. In what follows, I show that NEI is widespread and that it has been and is essential to the sciences.

Gambling Devices: Perhaps the most conspicuous success of NEI—though not the most important—is its use to infer the values of the probabilities associated with gambling setups, such as the tossed coin, the thrown die, the roulette wheel, and the urn full of colored balls. Without any more empirical information than that concerning the evident symmetries of these devices, we can infer exact probabilities for various outcomes. (For example, we infer a one half probability for a tossed coin's landing heads, a three fifths probability for drawing a red ball from an urn containing fifty balls, thirty of which are red, and so on.)

Furthermore, as noted above, if we are convinced that our information about the symmetry is correct (i.e., that there is not some hidden asymmetry, as in a loaded die), then we will cling to our inference about the probability even given vast amounts of unfavorable enumerative evidence. For example, if I am sure that a coin is perfectly balanced and tossed fairly, no run of heads will be long enough to force me to abandon my belief that the probability of heads is one half (see note 2).

Statistical Mechanics: The nineteenth century development of statistical mechanics relied on certain probabilistic assumptions about the movements of molecules, in particular, the assumption that in a given interval of time, all relevantly similar molecules are equally likely to be involved in a collision (rather as though the molecules that are to collide are determined by making selections from an urn). These assumptions, which allowed the derivation of the ideal gas law, various other parts of the kinetic theory of gases, and the second law of thermodynamics, were obviously not based on enumerative induction. No one was able to tabulate the frequencies of molecular collisions; on the contrary, the acceptance of the atomic theory of matter was largely brought about by the success of statistical mechanics. The probabilistic assumptions that provided the foundation for early statistical mechanics can only have been inferred by NEI. (For a historical sketch of the role of such assumptions in early statistical mechanics, see P. and T. Ehrenfest, 1959.)

The foundation (though not necessarily the truth) of these inferences has since been questioned (see Sklar 1994). But although physicists and philosophers have done their best to find other foundations for our knowledge of the probabilities of statistical mechanics, the fact remains that the distribution of those probabilities was first successfully inferred from symmetries. No further work on the founda-

tions of statistical mechanics can erase this fact; the best the NEI skeptic can do is to claim that the inference was a lucky accident.

Darwinian Explanation: Darwin's "fitness" is a probabilistic concept, measuring (roughly) the likelihood that an organism will survive and proliferate. (Thus survival of the fittest is not guaranteed: an unfit but lucky organism may be the one whose offspring inherit the earth. But it is overwhelmingly *probable* that the fitter organism will bear this responsibility. See Mills and Beatty 1979.)

Darwinian explanation of the predominance of some trait usually takes as a premise that that trait confers or conferred greater fitness than the alternatives.[5] This premise is a premise about relative values of probabilities. Because the species that bore the traits in question may be long extinct, enumerative tests of these probabilistic premises are not usually possible. The probabilities (or at least, their relative magnitudes) must be inferred by way of some kind of non-enumerative induction.

An example will show that this non-enumerative inference of probabilities is at least sometimes based on symmetries. Consider the claim that the fitness of large-beaked Galapagos finches increases (relative to that of other finches) as the number of large seeds increases (relative to smaller seeds). The reason for the relative increase in fitness is this: it is better for a small-beaked finch to find a small seed, because the small beak cannot crack a large seed, but it is better for a large-beaked finch to find a large seed, because large finches need the extra nutrition to maintain their large bodies. As the proportion of large seeds increases, the chance that the next seed a finch encounters will be large increases. Thus the probability of finding a favorable seed increases for large finches and decreases for small finches.

The argument for the claim about finch fitness, then, is based on a claim about probability: the more large seeds, the greater the probability of finding one. It is obvious to anyone that this claim is true. The reasoning involved is very similar to that concerning an urn full of balls (the urn is the island, the balls, seeds). As such, it is an instance of inferring probabilities from symmetries. Incidentally, in the case described, years of careful fieldwork have provided enumerative evidence that supports the non-enumerative inference made in a few seconds by the reader of this paper (Weiner 1994).

The successes just described strongly suggest that there is some robust procedure by which we go from symmetries to probabilities, and thus to frequencies, with considerable success. But what is that procedure, and why does it work?

Inference From Epistemic Symmetries

The Principle of Indifference may be seen as the consequence of putting the following constraint on the assignment of probabilities to a set of mutually exclusive outcomes:

> *(PI) Symmetries in the probability distribution ought to mirror symmetries in our knowledge concerning the outcomes.*

PI is not much help when we have an abundance of relevant information, such as a mass of statistics, for then there are no useful symmetries in our "knowledge space". The less we know, the more the rule comes into its own. When we have no information to distinguish two outcomes, it takes the familiar form: assign two possibilities concerning which we are equally ignorant, equal probabilities.

There is a standard objection to the Principle of Indifference. PI, it is claimed, cannot be phrased so as to pick out a unique probability distribution. The argument is usually presented as entailing the inconsistency of the principle, as follows: because many different probability distributions satisfy PI, the principle simultaneously recommends the ascription of conflicting probabilities. Thus it cannot be taken seriously as an inferential rule.

A well-known example that generates an inconsistent ascription of probability is that of the cube factory.[6] The factory produces cubes with edge length of up to two centimeters. What is the probability that the next cube produced has edge length of less than one centimeter? An answer: we know nothing about the edge length, except that it is less than 2 cm; thus, our probability distribution over the edge length should be distributed evenly over the interval between zero and two. It follows that the chance of the next cube having side length of less than one is one half. Another, inconsistent, answer: we know nothing about the cube's volume, except that it is less than 8 cm^3; thus, our probability distribution over the cube's volume should be distributed evenly over the interval between zero and eight. It follows that the chance of the next cube having a volume of less than one—that is, the chance that it will have a side length less than one—is one eighth.

It is possible to fend off arguments of this sort by imposing an extra constraint on the probability distribution, that it be invariant under certain transformations. A case can be made for this requirement by appealing to symmetries, which is certainly in the spirit of the original principle—indeed, in recent years, something of a philosophical arms race has developed in this area. The proponents of PI have suggested further rules for determining relevant symmetries, while their opponents have offered new counterexamples to those rules. (For a good summary, with references, see van Fraassen 1989, chap. 12.)

The entire dispute is, I believe, irrelevant to the problem posed at the beginning of this paper. The argument about the cube factory concerns the existence of a method for picking out unique probability distributions, and (less explicitly) the rationality of such a method, should it exist. In the case of the die and the other examples described in the last section, these issues simply do not arise.

First, we already know that, in the cases I have described, it is possible to arrive at a unique probability distribution. Upon viewing the dodecahedral die for the first time, we immediately see that the probability of any face landing uppermost is one in twelve. The problem is not whether this can be done, but how we do it.

Second, the question of the *rationality* of the one in twelve probability is quite secondary. What is striking is that the probability we choose is *correct*. It is this

fact that must be explained. By contrast, there is no such fact to be explained in the cube factory scenario, since we do not know the correct probability distribution over cubes.[7] (To gain such knowledge we would need to look inside the factory, at the very least.)

In short, there are two quite distinct problems. The first, exemplified by the case of the cube factory, is the problem of what probabilities, if any, can be rationally ascribed to events we know nothing about. (The most interesting consequence of solving this problem would be the justification of certain prior probability distributions for Bayesian inference; see Jeffreys 1939, Rosenkrantz 1981, and Jaynes 1983.) The second problem, exemplified by the case of the dodecahedral die, is the problem of how we successfully infer physical probabilities from symmetries. I have argued above that this inference must be based on knowledge, not ignorance. Physical probabilities are inferred from *physical* symmetries, that is, symmetries that are not in the mind but in the world.

Inference From Superficial Physical Symmetries

A proposal due to Henri Poincaré (1905) and adopted in a modified version by Hans Reichenbach (1949) founds NEI on physical symmetries. This proposal is similar in certain ways to my own, but it is inadequate, I will argue, because it takes for granted, without explaining, an important inference from symmetry to probability.

Poincaré's account of NEI is clearest in his treatment of the roulette wheel. The result of a game on the wheel, Poincaré assumes, is determined by the wheel's final position.[8] The position is represented by a real-valued variable θ ranging between 0 and 2π. Whether the wheel produces a red or a black number, then, is determined by θ. Call values of θ that correspond to red and black numbers respectively "red" and "black" values. Poincaré notes that, in virtue of the physical symmetries of the roulette wheel (or more exactly, of the way it is painted), red and black values of θ form intervals of equal size that alternate rapidly as θ goes from 0 to 2π. (See Figure 2, in which the shaded regions represent red values, the white regions black values, of θ.) From this symmetry—that is, from the physical symmetry of the pattern on the wheel—Poincaré claims we may infer a one half probability that a red number will be produced.

His reasoning is as follows. Let $d(\theta)$ be the physical probability distribution over θ, which is unknown to us. The probability of obtaining a red number will be given by summing the probabilities that θ falls into any particular "red" interval. More formally,

$$P(\text{red}) = \int_0^{2\pi} f(\theta)d(\theta)\cdot d\theta$$

where $f(\theta)$ is equal to one for "red" values of θ, and zero otherwise.

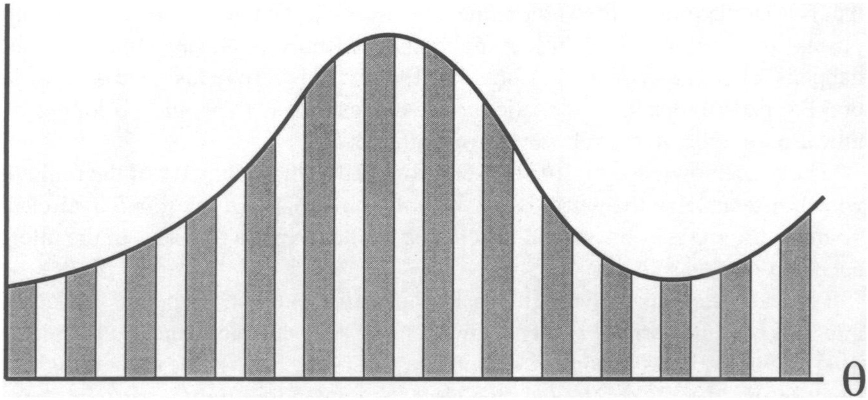

Figure 2. The probability of a red number is equal to the proportion of the area under $d(\theta)$ that is shaded. After Reichenbach 1949.

Now, provided that

1. The red and black intervals of θ are quite small, and so alternate rapidly, and
2. $d(\theta)$ is fairly smooth, in the sense that it does not fluctuate rapidly,

the above expression for the probability of red will be approximately equal to the ratio of red to black intervals, that is, one half. (More precisely, the ratio is one to one, which corresponds to a probability of one half.) This result holds because (a) a smooth $d(\theta)$ will remain approximately constant over any two neighboring (red and black) intervals, and (b) the ratio of any two such neighbors is the same (one half). We may be quite ignorant about the details of $d(\theta)$, then, but provided we know that it is fairly smooth, we can proceed from the symmetry of the roulette wheel's color scheme to a reliable estimate of the probability of a particular color.

The usual objection to symmetry-based approaches, applied to the Poincaré-Reichenbach approach, would be phrased as follows: why choose θ as the parameter with which to calculate the proportion of red to black outcomes, when there are many other parameters that would deliver different symmetries, and thus different probabilities? Poincaré and Reichenbach have a ready answer at their disposal: we choose θ because it is $d(\theta)$ that we know to be relatively smooth. (Thus the method for NEI is as follows: find an θ^* for which $d(\theta^*)$ is smooth, and let the ratio of red to black in $f(\theta^*)$ dictate the probability.[9])

The difficult question for Poincaré and Reichenbach concerns our knowledge of $d(\theta)$. How do we reach the conclusion that $d(\theta)$ is relatively smooth? The intuitive answer is that we expect $d(\theta)$ to be smooth because the roulette wheel has perfect circular symmetry. In fact, we do far better than this: from the symmetry of the wheel, we infer that the probability distribution over θ is *uniform*,

that is, that the probability is the same that any θ will be the wheel's final resting place. (To see that we are in fact inclined to reason in this way, consider what happens when the symmetry is broken. If the wheel were redesigned with little braking pads under the red sections, our estimate of $d(\theta)$ would no longer be uniform, or even, in the relevant sense, smooth.)

Thus our knowledge of $d(\theta)$ is itself grounded in the symmetry of the roulette wheel. It seems that the wheel is not the only thing that is going round in circles. To make progress, Poincaré and Reichenbach must explain the basis of the inference that $d(\theta)$ is smooth.

Poincaré attempts to avoid this task by pointing out that, as the red and black intervals become narrower and narrower and so alternate more and more rapidly, $d(\theta)$ needs to be less and less smooth. In the limit, he argues, it is enough to suppose that $d(\theta)$ is continuous, which is, he contends, surely a reasonable assumption. But "in the limit" here means "as the number of red and black sections on the wheel increases to infinity". Thus the assumption of continuity is relevant only to a wheel divided into infinitely many red and black sections. As Reichenbach sees, that is not the problem we face, and routinely solve, in real life. There are thirty-six red and black sections on the (simplified) wheel, no more.

Reichenbach does not attempt to solve the problem. He simply points out that the application of NEI to the wheel depends on an assumption about the smoothness of $d(\theta)$. He thus explicitly neglects, as Poincaré has implicitly neglected, the point at which NEI truly operates: our inference from the symmetry of the wheel to the smoothness of $d(\theta)$.[10]

Inference From Physical Symmetries in the Mechanism

It is my claim that NEI is based on symmetries in the mechanism of the chance setup in question. This position is, I think, intuitively appealing; in this section I will show that it is also correct.[11]

The first section shows why, physically, there is a correlation between certain symmetries and certain probabilities. The next two sections show how this physical fact serves as the basis for an account of NEI.

Roulette Once More

Consider the simplified roulette wheel. Assume for now that the laws of nature are deterministic, at least in the form in which they apply to the wheel. (I will later show that this assumption is not necessary.) It follows that the outcome of any game on the wheel is fully determined by some set of initial conditions. In any chance setup there will be several aspects of these conditions that change from game to game; I will call these aspects the *IC-variables*. In the case of the wheel the sole IC-variable is the speed with which the wheel is initially spun; call it ω. (Most setups will have more IC-variables; however, I will not deal with the multi-variable case in this paper. There are no serious complications (as shown in Strevens 1996).)

Suppose, then, that there is just one IC-variable ω. Let E designate the event of the wheel's producing a red number. As a consequence of the assumption of determinism, the laws of nature together with the mechanism of the wheel determine a function $U_E(\omega)$ which is equal to one just in case ω causes a red number to be obtained on the wheel, zero otherwise.

Observe that $U_E(\omega)$ has the same sort of symmetry as the function $f(\theta)$ introduced in the last section. "Red" and "black" values of ω form rapidly alternating bands; neighboring bands are of approximately equal size. (See Figure 3. Note that in $U_E(\omega)$, unlike $f(\theta)$, the red and black intervals slowly narrow. But they do so at the same rate, so that the ratio of a red interval to a neighboring black interval remains one to one.) The rapid alternation of $U_E(\omega)$ is a consequence of the fact that we can always change the outcome of a spin of the wheel by spinning a little faster or a little slower. The equal size of the black and red bands is a consequence of the physical symmetry of the wheel, in particular, of the fact that at any point in any spin, the wheel takes approximately equal time to rotate through a red segment as it does to rotate through a black segment. I will say that such a $U_E(\omega)$ is *microconstant* with ratio 0.5 (see the appendix for a formal definition of microconstancy).

Despite the similarity between $U_E(\omega)$ and $f(\theta)$, they are not to be confused. Whereas $U_E(\omega)$ expresses facts about the mechanism of the wheel, $f(\theta)$ expresses facts only about its paint scheme.

Now consider the probability distribution $D(\omega)$ over the IC-variable ω. (What I have to say is neutral with respect to the interpretation of $D(\omega)$. My only assumption is that there *is* such a distribution. If nothing else, we can always take $D(\omega)$ to represent frequencies of actual values of ω.[12]) As with $U_E(\omega)$ and $f(\theta)$,

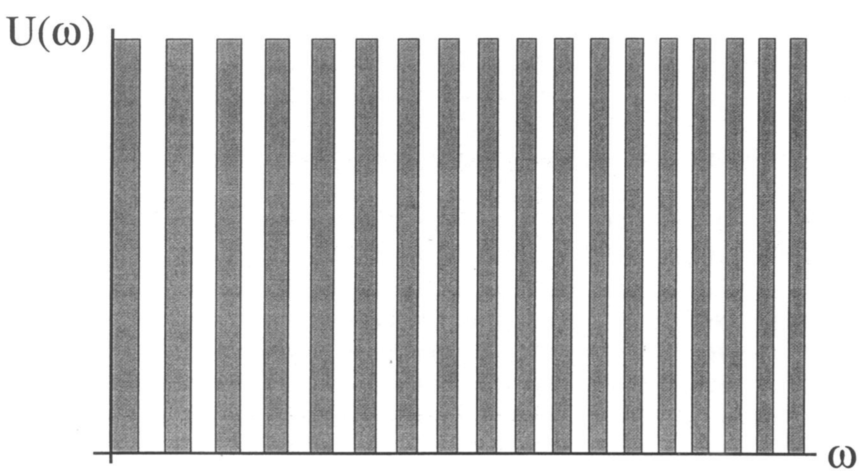

Figure 3. A portion of the function $U_E(\omega)$ (equal to one when ω produces a red outcome, otherwise zero).

there is an analogy between $D(\omega)$ and $d(\theta)$, but it is not exact: a value of θ describes an *outcome* of a game on the wheel, while a value of ω describes an *initial condition* of such a game.

The probability of the wheel's producing a red number can be derived from $D(\omega)$ and $U_E(\omega)$. It is simply the total probability of the "red" values of ω (i.e. those values of ω that result in a red number). $U_E(\omega)$ tells us which values of ω are red; $D(\omega)$ tells us how likely these values are. If E is the event of obtaining a red number,

$$P(\text{red}) = \int_a^b U_E(\omega)\cdot D(\omega)\cdot d\omega$$

where a and b are the minimum and maximum possible values of ω.

We can now apply Poincaré's purely mathematical observation to $U_E(\omega)$ and $D(\omega)$. For any sufficiently smooth $D(\omega)$, P(red) is—because of the geometry of $U_E(\omega)$—approximately equal to one half. Thus, provided that we know that $D(\omega)$ is reasonably smooth, we need know nothing more to infer a probability for E of one half. We can use the microconstancy of $U_E(\omega)$ to parlay some very general (probabilistic) information about the initial condition ω into an exact value for the probability of a red number.[13]

The same sort of treatment can be applied to any of the other familiar gambling devices (the die, the urn full of balls, a tossed coin, etc.). In each case, due to the symmetry of the apparatus, the function $U_E(\omega)$ is microconstant: it characterized by thin, evenly proportioned stripes (provided, of course, that E is one of the basic outcomes, such as a six being thrown or a particular ball being drawn). On the assumption that the relevant $D(\omega)$ is smooth, we can infer that the probability of the event E is $1/n$, where n is the number of basic outcomes. (A more formal, more general treatment is given in the appendix.)

This result provides a basis for NEI, as follows. Given a chance setup for which

(a) we have reason to think that the initial conditions are distributed smoothly, and
(b) we can infer from the symmetries of the setup mechanism that $U_E(\omega)$ is microconstant, with constant ratio p,

we can infer that the probability of an E event is p.

What is the role of the symmetries in such an inference? The symmetries do not themselves tell us the values of probabilities. Rather, they allow us to convert a very vague, very general piece of knowledge about probabilities—that ω is smoothly distributed—into a very specific piece of information, an exact value for the probability of an E event.

In the next two sections I consider the circumstances under which (a) and (b) above hold true, that is, the circumstances in which successful NEI is possible.

Knowledge of the Distribution of Initial Conditions

NEI requires a probabilistic premise, namely, that the probability distribution over the relevant initial conditions is "smooth". Where could this information come from? How can we know that some initial condition distribution function $D(\omega)$ is smooth? How can we come to have knowledge of the probabilistic distribution of an IC-variable such as the speed with which a roulette wheel is spun? Knowledge of $D(\omega)$ must be supplied either by enumerative induction or by non-enumerative induction.[14]

In the case of the roulette wheel, some kind of enumerative induction is probably at work. We simply know from experience that human actions such as the twirl of a wheel produce results that are smoothly heaped around some average value. (The experience is often gained while trying to produce results that require a far more finely honed distribution.)

We might also be able to arrive at a *non-enumerative* inference to the effect that $D(\omega)$ is smooth, if we knew enough about human physiology that we had some idea about the causes of variation in spins. This inference would depend on the distribution of some set of physiological initial conditions, which would have to be inferred in turn. At some point some enumerative induction must be done to get non-enumerative induction off the ground.

This initial enumerative induction might be on a grand scale. Call the kinds of variables in terms of which we usually work our "standard" variables.[15] It seems to be the case that, for whatever reason, our standard variables are usually smoothly distributed. If we go ahead and generalize from this observation (by enumerative induction), we arrive at the conclusion that most standard variable distributions are smooth. We may consequently take ourselves to have empirical grounds for adopting a revised and differently deployed "Principle of Insufficient Reason" of the following form:

> *In the absence of any reason to think otherwise, assume that any standard variable is fairly smoothly distributed.*

Let me stress that I am not proposing that our "standard" variables have any special logical status. They are simply the variables with which we prefer to operate, and which are, conveniently for us, for the most part smoothly distributed.

Note that this principle of insufficient reason is defeasible. Sometimes we have sufficient reason to think that a particular standard variable θ is *not* smoothly distributed. Consider a θ determined by a mechanism that, given a smoothly distributed $D(\omega)$ as input, produces a ruggedly distributed probability distribution $P(\theta)$ over its output. An example is my roulette wheel with brakes. We clearly ought not to assume that such a θ is smooth.

Knowledge of Microconstancy

The second premise required for NEI is that the relevant $U_E(\omega)$ is microconstant with some constant ratio p. Such information must come from knowledge of the

probabilistic mechanism in question together with knowledge of the laws of nature. There arises the issue of exactly which aspects of the mechanism and laws are relevant, a question made more interesting by the fact that we seem to get by with a small portion of the complete information. (Physics failures may be gambling adepts.)

What is necessary is whatever knowledge guarantees the existence of relevant symmetries in the operation of the mechanism, for example, whatever entails that a spinning coin takes about the same time for each half-revolution, or that a spinning roulette wheel takes about the same time for each 1/36th of a revolution. Such knowledge comes in two complementary parts: (a) what the laws of motion care about (distribution of mass, brake pads), and (b) what the laws do *not* care about (the color of a segment of the wheel, or the image on one side of a coin). The use of these facts to infer microconstancy does not require us to attend to the exact form of the dynamical laws of nature. Indeed, the facts remain the same (for the most familiar mechanisms at least) whether the laws of nature are Aristotelian, Newtonian, those of medieval impetus theory, or those of quantum mechanics. Thus the non-physicists' evident skill at pinpointing certain probabilities. (This observation has a further consequence: in so far as the laws of, say, evolutionary biology are based on probabilities founded in microconstancy, those laws do not depend on the details of the laws of physics, but only certain very general symmetries of the laws. See Strevens 1996 for the implications of microconstancy for reduction.)

That microconstancy is found in common gambling devices is reasonably clear. What, however, of ecosystems? What mechanism do we examine in order to divine a probability, and what symmetries do we look for? I have answered these questions in some detail elsewhere (Strevens 1996). The method, very roughly, is to regard the entire ecosystem as the mechanism, and then to ask which things affect, and which things fail to affect, the movement of the mechanism's parts—the falling seeds, the foraging finches, and so on. Exact probabilities may not be forthcoming, but qualitative results (concerning, for example, the relative advantages of large and small beaks) are within reach.

Finally, how can we apply this method of NEI to chance setups involving indeterministic mechanisms? In such cases, the effects of the indeterminism can be represented by an extra IC-variable ψ, representing the "error" caused by quantum fluctuations. This leaves us with a function mapping initial conditions to outcomes that is, in effect, deterministic. We may examine the symmetries of the mechanism as before, and, provided that $D(\psi)$ is sufficiently smooth, infer a corresponding probability.

Conclusion

Due to the microconstancy of certain mechanisms, and the smoothness of the distributions of certain variables, there is a correlation, in our world, between

symmetry and probability, and thus (in a world where frequencies usually reflect probabilities) between symmetry and frequency.

Our species has evidently picked up on this correlation, whether through learning or natural selection, and has exploited it for some time. Yet the reason for the correlation has not been clearly understood. It is because of this lack of understanding, I suggest, that the Principle of Indifference has been able to exert its hold. In what became the paradigmatic probabilistic experiments—those involving gambling devices such as the tossed coin and die—there is a direct correspondence between the symmetries in the mechanism and the symmetries in our knowledge about outcomes. Previous thinkers correctly saw that symmetries lead to reasonable beliefs about probability, but in framing the rule that enshrines this sort of inference, they wrongly took the epistemic symmetries, not the physical symmetries, as the proper source of these beliefs. In those cases where the beliefs turned out to be *true*, it appeared that NEI provided a priori knowledge of probabilities.

The truth about NEI is not as glamorous as that. Probabilities are not conjured from nothing, but are introduced in the premises of the inference in the form of an assumption about the distribution of initial conditions. Furthermore, empirical knowledge plays, if anything, a greater role in careful NEI than in enumerative statistical induction, for NEI is founded not just on knowledge of initial condition distributions, but also on fallible, empirical knowledge of the physical constitution of the relevant mechanism and of the pertinent laws of nature.

Yet knowledge of probabilities inferred from symmetries is knowledge acquired with little or no experiment in the traditional sense. It may not be a priori, but it is not experimental, either. As such, the further philosophical and psychological study of NEI has the potential to cast light on those sciences—such as statistical mechanics, evolutionary biology, and certain varieties of social science—in which experiment has played only a subsidiary role.[16]

Appendix

I provide here a more precise characterization of the mathematical properties of a function $\mathrm{U_E}(\zeta)$ that make NEI possible. For further details, as well as a discussion of a realistic case (which requires a method for dealing with multiple IC-variables), see Strevens 1996.

The *U-function ratio* of a U-function $\mathrm{U_E}(\zeta)$ over an interval $[x, y]$ of ζ is the proportion of the interval for which $\mathrm{U_E}(\zeta)$ is equal to one.

A U-function $\mathrm{U_E}(\zeta)$'s *constant ratio index* is determined as follows. Consider a partition of the range of ζ into intervals $[x, y]$. Call such a partition a constant ratio partition if the U-function ratio for each $[x,y]$ is more or less the same. Call the size of the largest interval in such a partition the partition's constant ratio index (CRI). Take the constant ratio partition of ζ that yields the smallest CRI q. That is the *U-function's* constant ratio index. (If $\mathrm{U_E}(\zeta)$ is constant, its CRI is zero.) If a U-function has a CRI of q, then any segment of $\mathrm{U_E}(\zeta)$ much longer than q will have the same U-function ratio. In such a case it makes sense to talk of the U-function ratio of the entire U-function.

A U-function is *microconstant* if it has a CRI that is very small.

Theorem: If a U-function is microconstant, with U-function ratio equal to p, then for any $D(\zeta)$ that is relatively smooth (that is, for any $D(\zeta)$ nearly constant over most small intervals of ζ), the probability of E is approximately equal to p.

Notes

[1]For early (eighteenth century) examples of the use of a principle of indifference to calculate probabilities concerning celestial objects, see van Fraassen 1989, 297–8. Laplace is among those cited; however, in his more philosophical writings, he seems to have realized that, inasmuch as probabilities generated by indifference are determined by ignorance alone, they may not correspond to real physical probabilities. Exegesis is difficult because Laplace's determinism inclined him against the existence of physical probabilities, and also because (perhaps for this very reason) he often failed to clearly distinguish physical and epistemic probability (see Hacking 1975, 131–3 for a discussion of the ways that Laplace did and did not make this distinction).

More recently E. T. Jaynes (1983) has applied a principle of indifference to determine probabilities in chance setups. In his most philosophical moments, Jaynes, like Laplace, admits that probabilities determined by ignorance are not necessarily the correct physical probabilities. (They are correct only if the problem is "well posed", but knowing that the problem is well posed requires some kind of real knowledge about the situation.)

These admissions of the epistemological limitations of the indifference principle only make it the more puzzling that our symmetry-based estimates of probability are so often correct.

[2]Of course, we are not normally absolutely certain about symmetries. A frequency of red faces significantly different from one twelfth will erode our belief in the symmetry of the die. But this is equally good evidence that we are sure of the link between probability and symmetry. In any case, the following more general claim seems correct: our subjective probability that the physical probability of a red face is 1/12 is never lower than our subjective probability that the die is symmetrical.

[3]"Rational probabilities" can mean rational subjective probabilities (as it does for the objectivist Bayesian) or logical probabilities, in the sense of Keynes (1921) and Carnap (1950).

[4]There may be cases of non-enumerative statistical induction other than those in which we infer probabilities from symmetries. But I know of none.

[5]Sometimes the trait itself does not increase fitness, but is correlated with some other fitness-increasing trait.

[6]The example is taken from van Fraassen 1989. It is based on the work of Joseph Bertrand (1889).

[7]I take it that this claim is uncontroversial. Even proponents of the principle of indifference do not think that the principle gives us the correct physical probability distribution over cubes (see note 1).

[8]Not all of Poincaré's comments are true of real roulette wheels. It is perhaps best to interpret him as describing a simplified roulette wheel, in which there is a fixed pointer instead of a ball. When the wheel comes to rest, the number next to the pointer wins.

[9]Note that if θ^* is a parameter for which red and black intervals are not equal, then $d(\theta)$ and $d(\theta^*)$ cannot both be "fairly smooth" in the required sense.

[10]It is worth pointing out a further problem with the Poincaré-Reichenbach proposal, closely related to that just described. The proposal depends on there being two probability distributions that each describe the outcome of the same trial, one of which is continuous (for the wheel, θ). In many important cases, there is no such continuous variable. What continuous variable, for example, describes the outcome of a coin toss, or a drawing from an urn? Or a spin of a non-simplified roulette wheel in which each colored section has a hollow to capture the ball?

[11]The approach taken in this section is similar in some respects to the "method of arbitrary functions" that emerged from Poincaré's work, especially as developed by Eberhard Hopf (Hopf 1936; for summaries in English see von Plato 1983 and Engel 1992).

Hopf examines, as I do, the behavior of the function mapping initial conditions onto outcomes, and shows that for some gambling devices, Poincaré's insight can be applied. However, like Poincaré, Hopf is interested only in the behavior of these functions as certain parameters (the number of red and black sections or the initial velocity of the wheel ω) approach infinity. Since these parameters do not, in fact, approach infinity, the physical significance of Hopf's work is at best unclear.

Eduardo Engel has written a monograph extending Hopf's work in many ways (Engel 1992). The chief mathematical technique is, as in Hopf, to let parameters approach infinity. But Engel also arrives at some results concerning the *rate of convergence* on the probability as the parameters increase (for certain kinds of initial condition distribution functions). These results do have physical significance. His section on coin tosses is a case in point: he shows that, for realistic parameters and a sufficiently smooth initial condition distribution function, the probability of heads will be very close to one half. This work has a rather different flavor from mine; there is much talk of convergence (a property of the sequences generated when a parameter goes to infinity), but nowhere is there a discussion of what I call microconstancy (a property of real physical systems in which parameters are fixed). Also, Hopf and Engel are not concerned with the epistemological problem addressed here. They seek to explain why the physical probability of a red value is one half, not how we come to know this fact.

[12]A frequency interpretation of the initial condition distribution function introduces a technical difficulty dealt with in Strevens 1996.

[13]If you are inclined to doubt that croupiers' functions $D(\omega)$ are smooth, consider this: if the functions are not smooth in the relevant sense, it would very likely be the case that different croupiers would turn up red numbers with different frequencies. Casinos have spent much time and money ensuring that this is not the case; thus, for gambling devices at least, it is reasonable to suppose that the relevant functions $D(\omega)$ are smooth. See the next section for further discussion.

[14]Leonard Savage (1973) has suggested, in connection with Poincaré's explanation of NEI, that a distribution such as $D(\omega)$ can be interpreted as a subjective probability distribution. But while this might explain the fact that we agree on a probability for red (because we all have a smooth subjective probability distribution over ω), it cannot explain the objective correctness of the agreed-on answer (nor did Savage mean it to do so).

[15]One way of characterizing standard variables is as follows: they are those variables that induce measures directly proportional to the SI units. Thus inches are standard measures of length, while inches* are not, where an object's length x^* in inches* is derived from its length x in inches as follows: $x^* = \sin x + x$.

[16]Thanks to Barry Loewer and Tim Maudlin for helpful comments.

References

Bertrand, J. (1889) *Calcul des Probabilités*, Paris: Gauthier-Villars.

Carnap, R. (1950) *The Logical Foundations of Probability*, Chicago: University of Chicago Press.

Ehrenfest, P. and T. (1959/1912) *The Conceptual Foundations of the Statistical Approach in Mechanics*, translated by M. Moravcsik, reprint, New York: Dover Publications.

Engel, E. (1992) *A Road to Randomness in Physical Systems*, *Lecture Notes in Statistics vol. 71*, Berlin: Springer-Verlag.

Hacking, I. (1975) *The Emergence Of Probability*, Cambridge: Cambridge University Press.

Hopf, E. (1936) "Uber die Bedeutung der willkürlichen Funktionen für die Wahrscheinlichkeitstheorie", *Jahresbericht der Deutschen Mathematiker-Vereinigung* 46, I. Abt., 179–95.

Jaynes, E. T. (1983) *Papers On Probability, Statistics, And Statistical Physics*, edited by Roger Rosenkrantz, Dordrecht: Reidel.

Jeffreys, H. (1939) *Theory of Probability*, Oxford: Clarendon Press.

Keynes, J. M. (1921) *A Treatise on Probability*, London: Macmillan.

Mills, S, and J. Beatty. (1979) "The propensity interpretation of fitness", *Philosophy of Science* 46:263–286.

Von Plato, J. (1983) "The method of arbitrary functions", *British Journal for the Philosophy of Science* 34:37–47.

Poincaré, H. (1905) *Science and Hypothesis*, reprint, New York: Dover Publications.

Reichenbach, H. (1949) *The Theory of Probability*, second English edition, Berkeley: The University of California Press.

Rosenkrantz, R. (1981) *Foundations and Applications of Inductive Probability*, Atascadero: Ridgeview.

Savage, L. J. (1973) "Probability in science: A personalistic account", in P. Suppes et al. eds. *Logic, Methodology and Philosophy of Science IV*, North Holland: Amsterdam, 467–483.

Sklar, L. (1993) *Physics and Chance*, Cambridge: Cambridge University Press.

Strevens, M. (1996) "The Probabilistic Structure of Complex Systems", Ph.D. dissertation, Rutgers University.

van Fraassen, B. (1989) *Laws and Symmetries*, Oxford: Oxford University Press.

Weiner, J. (1994) *The Beak of the Finch*, New York: Knopf.