Symmetry and Probability*

Jill North†

October 13, 2006

1. Introduction

There is a strong connection, in our world, between symmetry and probability. We often rely on symmetries to infer the probabilities of different possible outcomes. Based on the physical symmetry of a die, we infer that the probability of getting any particular face when I throw the die is one in six. The symmetry of the die also tells us to have the corresponding degree of belief: our degree of confidence in any given face is also one in six. When I throw the die a bunch of times, we do observe each face approximately $\frac{1}{6}$th of the time.

It seems as though we can infer these probabilities just by looking at the symmetry in the shape of the die, before witnessing any actual throws. We see that there are six different, completely symmetric ways the die could land, and we distribute the probabilities, both the objective chances and our degrees of belief, evenly among them.

We often make this symmetry-based inference to probabilities. And these inferences tend to succeed: the probabilities approximate the observed frequencies. It seems as though the symmetries in a situation, where there

*For helpful comments and discussion on earlier versions or parts of this paper, I am grateful to David Albert, Hartry Field, Michael Strevens, Ted Sider, the Corridor group, the students in a seminar at NYU in the fall of 2005, especially Matthew Kotzen and Jonathan Simon, and audience members at the University of Maryland, College Park and the Bellingham Summer Philosophy Conference at Western Washington University in 2006. I am especially grateful to Branden Fitelson, who commented on a version of this paper at Bellingham, and also gave me extremely fruitful feedback via email.

†Dept. of Philosophy, Yale University: jill.north@yale.edu

are any, indicate the probabilities of the different possible outcomes. These symmetries indicate the probabilities even before we have witnessed any actual outcomes.

This has led to a tradition of maintaining that this a priori symmetry reasoning suffices for inferring probabilities. On this view, a priori symmetry considerations based on our epistemic situation correctly indicate the objective chances of outcomes. They also dictate what degrees of belief we should have in their obtaining.

I don't think the kind of a priori principle that is usually cited is what explains the success of our probability assignments, or what justifies the probability assignments we make. I argue that these symmetries do not indicate the objective chances of outcomes; nor do they dictate reasonable credences about the chances. Where symmetry considerations do succeed, they are not the a priori ones people have taken them to be. I propose to explain these inferences on the basis of empirical symmetries in the structure of our world.

2. Symmetry principles

First let me explain a bit more what I mean by our use of symmetries, or a symmetry principle, for inferring probabilities of different possible outcomes.

Suppose I have a die I am about to throw. I have no information that the die is loaded, no reason to suspect it is not a fair die. There are six different ways the die could land, with any one of its six sides facing up. What probability should I assign to each possible one?

It seems the reasonable thing to do – indeed, the *only* reasonable thing to do – is assign the *same* probability to each of the six possibilities, and conclude that each face has a $\frac{1}{6}$ probability of coming up. This seems like the right physical probability, or objective likelihood, of any getting any particular face on the die, as well the subjective probability, or degree of belief, I should have in each one.

(You might think there are no such chances here. You might have the view that if the laws are deterministic, there can be no objective chances other than 0 or 1. Or you might think there are no objective probabilities at all in the world. I disagree with both of these views, but wish to leave this aside here. If you think there are no objective probabilities, then you can translate (some of) this talk of chances into subjectivist terms. (It is not immediately obvious how much of what I say can be accommodated by the subjectivist. I take it the subjectivist will have some way of making sense of our ordinary chance talk, i.e. why it seems as though there are chances in the world, and what are the things that physicists call 'chances'.) I continue on the assumptions that there are objective probabilities, and that these can differ from 0 or 1 even if the world is deterministic. (If objective probabilities simply correspond to robust patterns in the actual relative frequencies of outcomes in the world, then they can be different from 0 or 1 regardless of determinism.) I think my conclusions remain either way, but leave this to the reader to decide.)

Back to the die. Since I have no information that the die is loaded, and since there appears to be no other relevant asymmetry which would make a difference to the outcome of a throw, if I were to say that the face with number 5 on it, say, has a chance that is greater than $\frac{1}{6}$, this would strike you as a completely arbitrary, unreasonable preference; you would think me irrational for betting at different odds. My weighing one face more heavily in the absence of any positive reason for doing so is to base my probability inference on factors that have nothing to do with the actual state or behavior of the die.

Instead, I should conclude that each face in fact has a $\frac{1}{6}$ chance of coming up; if I am rational, my degrees of belief should follow suit. And the reason for this inference to symmetrically distributed probabilities – the reason I assign the *same* probability to each possibility – is the corresponding symmetry in the die, and my epistemic state with respect to it (the fact that I have no extra information that the die is biased in some way).

In order to figure out the probabilities of different possible outcomes, we tend to reason as follows. First, look at the symmetries in the situation. Next, use these to determine the outcome space, the set of elementary or fundamental possibilities to which we are going to attach probabilities. (The symmetry of the die tells us the outcome space is the set of integers $\{1,2,3,4,5,6\}$.) Finally, assign an equal probability to each such possibility. Distribute the probabilities, both the objective chances and our degrees of belief, uniformly over the possibility space. Calculate the probabilities of any non-elementary events from these basic probabilities.

(Michael Strevens (1998) calls this type of inference a 'non-enumerative statistical induction'. We don't calculate the probabilities of different outcomes by tallying up the frequencies with which we observe them to occur; we simply examine the relevant symmetries in the set-up.)

In figuring out the probabilities of outcomes, we rely on this principle: assign an equal probability to each basic possibility, where what are the possibilities is determined by the symmetries relative to our epistemic situation. This is often called an 'indifference principle', since it tells us to infer a probability distribution that is indifferent among the possibilities which, for all we know, could obtain.

We do rely on this kind of principle in our everyday reasoning about probabilities. What is more, this principle seems to *work*. The probabilities turn out to approximate the (long-run) relative frequencies. In our experience, a fair die does tend to land on number 5 approximately $\frac{1}{6}$th of the time, an unbiased coin does land heads up in approximately half the tosses.

This is remarkable. Surprisingly little information is needed to determine the correct probabilities of outcomes. Just looking at the relevant symmetries in a situation, before we have witnessed any actual outcomes, allows us to make successful probabilistic predictions. There is a distinguished tradition, from Laplace[1] and continuing on in different guises in the work of physicists and philosophers such as E. T. Jaynes,[2] of maintaining that an a priori indifference principle suffices to indicate what the chances are, and to dictate what degrees of belief we should have in those chances. On this view, we can have a priori knowledge about the probabilities of possible outcomes.

But is this reasoning justified? Do the apparent symmetries suffice to indicate the probabilities of outcomes, even before we observe any actual outcomes? If they do not, then what explains the success of our everyday inferences from symmetries to probabilities? I first consider whether symmetries suffice for inferring the objective probabilities of outcomes. I then turn to symmetry and subjective probability.

[1]See (van Fraassen, 1989, ch. 12) for discussion of historical uses of indifference for calculating probabilities.

[2]See Jaynes (1983), in particular the papers "The Well-Posed Problem" and "Prior Probabilities."

3. Physical probability

Consider the chances in our theories of physics. Do symmetries suffice to indicate these physical probabilities? One example suggests they do: the probability distribution of classical statistical mechanics.[3] Certainly the physics books cite a symmetry principle in justifying this distribution.[4] Though I agree that symmetries play a role, I think the books are wrong about what kind of symmetries these are.

To see this, we need to understand a bit about the theory's statespace. In statistical mechanics, the space of fundamental physical possibilities, the fundamental statespace, is called phase space. This is a mathematical space that represents all the possible physical states of a system. In classical mechanics, a system's fundamental state is given by the position and momentum of each of its particles. The phase space of a classical n-particle system has $6n$ dimensions, one for each of the three position and momentum coordinates for each particle (for particles free to move around in three spatial dimensions).

Each point in a system's phase space picks out an exact possible total state, or microstate, of the system. This is its most precisely specified state, given by the position and momentum of each of its particles. A system's macrostate is specified by its macroscopic features (things like average temperature, pressure, and volume). In general, corresponding to a given macrostate, there are many different possible microstates, many different ways the system's particles can be arranged so as to give rise to the same macroscopic features. A macrostate corresponds to a region of phase space, each point of which picks out a microstate realizing that macrostate.

Phase space is the fundamental possibility space of statistical mechanics. In order to make predictions about a given system, the theory places a probability distribution over its phase space. The probability distribution used by

---

[3] I stick to classical statistical mechanics. I think my conclusions should apply to the distribution of quantum statistical mechanics, but this raises technical issues not central here.

[4] See e.g. Landau and Lifshitz (1980), Tolman (1979). To be fair, Tolman goes on to say that the choice of this distribution "can be ultimately justified only by the correspondence between the conclusions which it permits and the regularities in the behaviour of actual systems which are empirically found" (Tolman, 1979, 59); he nonetheless takes a priori indifference to be a prima facie, if fallible, reason for the uniformity assumption.

statistical mechanics is the uniform one. (The distribution is uniform with respect to the standard Lebesgue measure defined over the canonical position and momentum coordinates, the Liouville volume measure. This is the familiar way of calculating volume, applied to higher-dimensional spaces like phase space. To find the volume of a region, multiply its extensions along each dimension in the space. The uniform distribution with respect to this measure says that the probability of a region is proportional to its volume calculated in this way.) This probability distribution says that each microstate compatible with a given macrostate is equally likely. Statistical mechanics thus assigns the same probability to each fundamental possibility.[5]

There is no question that these probabilities are empirically successful. Among other things, statistical mechanics is used to explain thermodynamic phenomena, such as entropy increase in isolated systems. A key component of this explanation is the fact that, according to statistical mechanics, higher entropy states are much more probable than lower entropy ones.

But what justifies the theory's assumption of a uniform distribution over phase space? What makes this the correct assignment of probabilities to microstates? Since the set of microstates compatible with a given macrostate is continuously infinite (position and momentum take on continuum many values), there are many different, mathematically legitimate probability assignments we could use. (I return to this in a moment.) Why this particular one? There is something perhaps more generally puzzling about these probabilities. Any given system, at any time, will be in some exact microstate or other; a classical system's particles always have precise position and momentum values. What justifies our assigning probabilities to the different microstates a system could be in, if one always in fact obtains? Why use probabilities at all in the theory of these systems? This is especially puzzling if we consider the universe as a whole, as one big statistical mechanical system: if there is only one system, it seems we can't appeal to relative frequencies to understand these probabilities.

[5]More accurately, statistical mechanics assigns the same probability to equal phase space volumes on the standard measure; this a probability density function, which is the derivative of the probability distribution. For the purpose of this paper, I will continue to talk about the probabilities assigned to fundamental possibilities, represented by points in phase space, though this really should be understood as probability densities.

This is where a symmetry or indifference principle is supposed to come in. Physics books begin by noting that, although any system is always in one particular microstate or other, we don't know *which* one. We could not possibly know which microstate actually obtains. That would require our being able to measure the exact position and velocity of each particle, which is something that, for typical macroscopic systems, we cannot in general do.

That's the first thing the books point out: there is a continuous infinity of microstates consistent with our macroscopic knowledge. The next thing they note is that we have no reason to think a system is any more likely to be in one possible microstate rather than any other. They conclude that we should assign an equal probability to each one. They suggest the only thing we *can* do is assign the same probability to each possible microstate.

This is similar to the reasoning we used for the die. Just as for the die, here we are told that in the absence of any reason to the contrary, we should weight each possibility equally. Once again, it is the symmetry relative to our epistemic situation – in this case, the fact that we are limited to knowledge about a system's macrostate – that determines the set of elementary possibilities to which we initially assign probabilities. Hence the 'hypothesis of equal a priori probabilities' in the physics textbooks: we are supposed to assume, even before having any evidence about a system's behavior, that each microstate compatible with its macrostate is equally likely. The books suggest we can determine the correct physical probabilities on the basis of a priori symmetry or indifference considerations, just as we seem able to do in our everyday reasoning about the outcomes of die throws and coin tosses.

There are two general reasons the probabilities of statistical mechanics can't be inferred on the basis of this kind of indifference principle. The first is familiar, but worth spelling out. The problem is this: indifference alone won't yield a unique probability assignment. An indifference principle will assign different probabilities depending on the parameters with which we describe the possibilities. And, in general, there is no a priori way of picking out a unique and non-arbitrary set of such parameters.

The familiar example of the cube factory illustrates this problem.[6] A factory produces cubes of side length $\leqslant 1$ foot. What is the probability that a cube produced by this factory will have side length $\leqslant \frac{1}{2}$ foot? The answer

[6]See van Fraassen (1989, 303).

depends on the parameters we use to describe the possibilities. We get different answers depending on whether we distribute probabilities uniformly over cube length (probability$\frac{1}{2}$), side area ($\frac{1}{4}$), or volume ($\frac{1}{8}$).[7] In general, any probability assignment depends on the choice of description of a situation. With different parameters, we get different probabilities assigned to the various possibilities.

The problem is that the principle of indifference tells us, *given* a particular parameter space (and our symmetric information with respect to it), we must distribute probabilities evenly over it. It does not also say *which* parameter space we must use. So whenever there are different, equivalent ways of describing the possibilities, there are different "uniform" probability assignments we could make. Even for the die, there are different ways of describing the possibilities, depending on the type of outcome we are interested in; consider the probability of its landing on an even as opposed to an odd number, a prime versus a non-prime, or a one versus a non-one.[8] (If we stipulate that we're interested in the probability of one of the six faces, there will be a unique possibility space.) Indifference does not say how we must carve up the possibilities before assigning probabilities to them; yet different carvings yield different probabilities. Further, if there are different parameters that are non-linearly related to one another, as for the cubes, then indifference will yield incompatible probability assignments. And since there is no a priori reason for thinking that any of one of these

[7] First assume that each possible edge length is equally likely. Then we get the answer $\frac{1}{2}$: out of all the 1-foot-edge-length cubes produced by the factory, about half of them will have edge length $\leqslant \frac{1}{2}$ foot. Next distribute probabilities uniformly over the area of a cube's side. Then we begin with the fact that all cubes have side area $\leqslant$ 1ft^2. Assuming that any cube meeting this condition is equally likely, we get that $\frac{1}{4}$ of them should have a side length $\leqslant \frac{1}{2}$ ft, that is, side area $\leqslant \frac{1}{4}$ ft^2. Finally, start with a cube's volume. Then all cubes produced by the factory have a volume $\leqslant$ 1 ft^3. Out of these, the cubes with sides of length $\leqslant \frac{1}{2}$ ft are the ones with volumes $\leqslant \frac{1}{8}$ ft^3. Now we get the answer $\frac{1}{8}$.

[8] The last example is from Sklar (1993, 199). Sklar notes that if we impose the constraint that the elementary possibilities correspond to "indecomposable" events, then indifference will yield unique probabilities; and that this is an important difference from the infinite case, for which the indecomposable events (the points) all have probability zero, on any distribution. But note that we are still left with the conclusion that indifference won't suffice, even in the finite case, since (a) we need an independent notion of indecomposability, and (b) which events are indecomposable in this sense cannot be known a priori.

equivalent descriptions is the "correct" one, any description we do choose will seem arbitrary. But then the resultant probability assignment will seem just as arbitrary.

Note that this is precisely what proponents of indifference claim it can do for us. Symmetries are supposed to suffice to tell us what the correct probabilities are whenever we have no other information to go by. Symmetry tells us to assign probability $\frac{1}{6}$ to each face before tossing a die, probability $\frac{1}{2}$ to each side before tossing a coin. Of course, knowing more about the cube factory could tell us what the right probabilities are: we could infer them from the frequencies with which different-sized cubes are actually produced. This reinforces the conclusion that a priori symmetry considerations on their own cannot do the job.

The problem only worsens for infinite possibility spaces, especially uncountably infinite spaces like phase space. Then there will be infinitely many ways of assigning the same probability to each possibility.

Let's spell this out. Start with a countable set of possibilities, and try to distribute probabilities uniformly over it. We can't assign each possibility an equal finite probability, since the probability of all the possibilities would be greater than one, in violation of a probability axiom. We could assign each possibility zero probability. This will violate countable additivity: the probability of the entire space is 1; the sum of the individual probabilities is zero. (We could assign each possibility infinitesimal probability; this too violates countable additivity, since the sum of a countable set of infinitesimals will not converge to 1.[9])

In order to take a uniform distribution over a countable space of possibilities, then, we will be forced to violate countable additivity. Although countable additivity seems intuitive, it is controversial whether to include it as a probability axiom. Yet even if we are willing to drop it, there remains the deeper problem that there is no unique way of assigning each possibility equal probability. There will be infinitely many ways of satisfying our con-

[9] The sum of a countably infinite sequence is the number that larger and larger finite partial sums get closer and closer to, if there is such a number; if there is not, we say the series diverges. There is no such number for a countable sequence of equal-sized infinitesimals. Even setting this aside, clearly a countable sequence of infinitesimals will not converge to 1 (if it converges, the sum will be infinitesimal), and this rules out a uniform distribution; I thank Hartry Field for this last point.

straints, that each possibility get the same probability, and the whole space gets probability 1.

Turn to an uncountable space in order to sidestep the issue of countable additivity. Take the points on the real line between 0 and 1. Consider a distribution that assigns probability $\frac{1}{2}$ to the interval $(0, \frac{1}{2})$ and probability $\frac{1}{2}$ to the interval $(\frac{1}{2}, 1)$. Now consider a different assignment: probability $\frac{1}{4}$ to the first interval, $\frac{3}{4}$ to the second. Each distribution assigns an equal, zero, probability to each possibility (or point) in the space, and probability 1 to the entire space. Hence each is a legitimate way of assigning uniform probabilities. Yet since there is no a priori reason to think that either one is the "correct" uniform distribution, indifference won't tell us what probabilities to infer.

The distinctive problem that arises for infinite spaces is that the probabilities of the individual possibilities don't determine how the probabilities must be distributed over infinite (measurable) sets of them. The probabilities of the elementary possibilities (the points) on the real unit interval don't suffice to give the probabilities of sub-intervals of the space (specifically, sub-intervals of the form (r_1, r_2), r_1 and r_2 two distinct reals between 0 and 1). Indifference dictates that the probability of each elementary possibility must be zero (or infinitesimal). But this requirement can be satisfied by many different distributions, relative to different parameterizations. Again take the real unit interval. Consider two coordinatizations of the points – say, by x and x^2 – and a uniform distribution over each one. Each of these distributions assigns an equal, zero, probability to each point in the space; and each assigns an equal probability to equal-sized sub-intervals (equal-sized according to its own paramaterization). The two distributions disagree on the probabilities of infinite sets of points. In particular, they disagree on the probabilities of the sub-intervals, since they disagree on the very sizes of these sets of points. Yet there is no a priori way of saying which is the "correct" coordinatization, or the "real" size of the intervals, and so the "correct" way of assigning uniform probabilities. There's no reason to think that one coordinatization is intrinsically better – that would presuppose we already knew what the right probabilities are. Once again, the conclusion is that indifference fails to yield unique probabilities. Here, indifference fails to tell us the probabilities of anything that gets a non-zero probability.[10]

(As a practical matter, we only ever seem faced with finitely many possibilities. This is because we are ignoring all the micro-possibilities, the infinitely many ways in which the macroscopic options can be realized by different microscopic situations.)

There is a second reason we can't infer the probabilities of statistical mechanics on the basis of an a priori indifference principle. We use these probabilities to predict and explain the frequencies with which actual outcomes occur. Simply put, the actual frequencies can diverge from the apparent symmetries. The frequency data could disconfirm any initial symmetric distribution. Since the observed frequencies must be some evidence of the actual physical chances, and since there's no *a priori* reason the actual frequencies must track the apparent symmetries, symmetries can't suffice to indicate the actual physical probabilities.

We don't think systems would behave any differently if we *did* know their exact states. This further suggests that symmetries with respect to our epistemic state, and probabilities based on it, cannot factor into scientific explanations of their behavior. Indeed, indifference would tell us to assign different probabilities if we *did* know systems' exact states. Although in that case we wouldn't need the uniform distribution (or probabilities at all) for making predictions, there is no reason to think the uniformity assumption would suddenly *stop* being successful if that were the case. Finally, indifference tells us to posit uniform probabilities in any world of which we are similarly ignorant. But consider a non-statistical mechanical world: this would yield the wrong predictions in a world like that.

You might respond that indifference is a prima facie, defeasible basis for inferring probabilities. Yet since there is no reason an initially uniform assignment is any more likely to succeed than any other, what could justify the assumption of a uniform distribution, over a particular parameter space, to begin with? In the absence of evidence, there is no more reason to choose an initial symmetric distribution rather than some other. So choose one, and

[10]This is not to say that indifference does suffice to tell us the probabilities of each of the countably many elementary possibilities. Rather, even *given* the assumption of equal-probability points – which in an uncountable space can *only* be done by assigning them each zero or infinitesimal probability – no other probabilities are settled.

update the initial probabilities in the right way as the evidence comes in.[11]

(Hence even the Bayesian, who allows for the updating of initial probabilities on the basis of evidence, cannot justify an initial distribution on the basis of indifference. The problem remains: what could justify the symmetric distribution to begin with, before we conditionalize on the frequency evidence? Of course, the Bayesian will require *some* initial distribution over credences to be able to take that evidence into account. It is not that we are *un*justified in choosing any priors at all to get us going, but rather that without any evidence, an initial distribution that is not symmetric seems just as good as a symmetric one. I return to this in section 5.)

Thus, even if symmetries could yield a non-arbitrary, unambiguous assignment, this still wouldn't suffice to tell us the physical probabilities. Consider the case of a finite, discrete statespace, such as for the six faces of a die. Here there is a unique uniform distribution. Still, nature might not oblige and distribute systems this way.

4. Empirical justification

What then tells us to use the uniform distribution in statistical mechanics, if we have no other information to go by? Answer: we do have other information – empirical input from the world. We must rely to some extent on evidence of actual frequencies in order to confirm the probabilities we use in physics. And statistical mechanics, with its probability distribution, is an extremely successful empirical theory.

Now, the statistical facts on their own underdetermine the exact form of the probability distribution. Other, not-completely-uniform distributions should yield just as good probabilistic predictions. Why the uniform one? Answer: this distribution is uniform over the structure needed for the dynamics.

Let me explain. Recall the statespace of statistical mechanics. The elementary possibilities are given by the different possible combinations of mo-

[11] For the purpose of this paper, I am assuming that the right way to update is by means of something like Bayesian conditionalization, though I don't have any argument for this here. It suffices to assume that there is some correct method for taking in evidence, according to your preferred confirmation theory.

mentum and position coordinates for each of a system's particles. The uniform distribution counts each such possible combination as equally likely.

You might wonder why we use the momentum and position coordinates to describe the fundamental states in the first place. We could describe systems' states differently, using different coordinates. So we seem to be back at the problem that there is no unique, non-arbitrary set of parameters over which to distribute probabilities uniformly. Here, the question is why we are justified in making the uniformity assumption with respect to the position and momentum coordinates of phase space.

The answer is that these are the canonical coordinates, in terms of which there is a particularly simple formulation of the dynamics, Hamilton's equations. Hamiltonian dynamics requires a certain amount of structure, called symplectic structure. This is all the structure that is needed for the dynamics of classical systems.[12] And phase space has this structure. A probability distribution that is uniform over this space then requires no further structure in addition to what is already needed for the dynamics.

This then gives us two reasons for this distribution. (1) It yields empirically successful predictions. (2) It is uniform over the structure required by the dynamics; it is the simplest, most natural distribution, given the dynamics. It requires no further structure over and above the structure that is "already there" for the dynamics.

You might worry that the simplicity considerations which lead us to prefer Hamiltonian dynamics in the first place are somewhat a priori or arbitrary, since we can formulate the dynamics differently, in terms of different variables. Yet insofar as we think the simplicity and invariance of the dynamics are not arbitrary but track genuine features of the world, we can avoid this worry.[13]

[12] A symplectic manifold comprises a differentiable manifold and a symplectic two-form. The symplectic two-form is a geometric object that encodes information about the kinematics. We can think of the symplectic form as picking out the canonical coordinates, the coordinates such that transformations between them preserve the Hamiltonian dynamics. The Hamiltonian is a scalar function defined on the manifold, which encodes information about the dynamics of the system, including the forces acting on it. A symplectic manifold with a Hamiltonian then has all the necessary structure for the dynamics of classical systems. For a nice discussion of symplectic structure in classical mechanics, see Singer (2001).

[13] You might also worry about other formulations of classical dynamics, such as La-

Generally, in physics, we infer the existence of the simplest, minimal structure needed for the dynamics. We take this to be a feature of the dynamics and the underlying structure of the world, not the way in which we formulate the theory. We assume that if there were additional structure, it would show up in the dynamics. For example, from the theory of relativity, we infer that spacetime has no identity-of-spatial-location-across-time structure: the theory does not require this structure, and we correspondingly infer there is none. Likewise, the time-translation invariance of the dynamical laws suggests that time itself has no preferred-temporal-location structure. Invariances in the laws suggest the corresponding symmetries in the structure of the statespace of the theory, and of the world that theory describes; they suggest the lack of any structure that would be needed to support the corresponding *a*symmetry. These invariances are things we can check by inspecting the dynamical laws.

In statistical mechanics, a non-uniform probability distribution would require additional structure, a kind of "preferred-point (or -region) in phase space" structure. Since we do not need this structure for the dynamics of classical systems, and since we can formulate a successful statistical mechanics without it, we should infer a uniform distribution as the one that accurately reflects the underlying dynamical structure of the world.[14]

To put it another way, we tend to think the world must have at least the amount of structure that is required to formulate its fundamental dynamics. Although it might seem as though we should be able to formulate the dynamics using different variables, and have a statistical mechanics based on the structure of its statespace, we do know that symplectic structure suffices for the dynamics; and we do not know of any other structure that does. It seems to be a fact about our world that its dynamics requires this particular structure. Symplectic structure is what's invariant under transformations that preserve the dynamical laws. A uniform distribution is then the

grangian mechanics. For arguments that we have reason to prefer the Hamiltonian formulation and its structure, see my North (ms).

[14]This is akin to other inferences we make in physics. Modern geometric formulations of physics emphasize the distinction between the features ascribed to a space because of the particular coordinate system, and the structural features of the space itself. The invariance of coordinatizations under certain transformations, for instance, indicate structural symmetries of a space.

unique distribution requiring no additional structure. This inference is not conclusive; we cannot be certain that the inference to the minimal structure required by the dynamics is correct. But it does seem to be a reasonable inference, and is one we generally make in physics.

The justification for the uniform distribution then bottoms out at the contingent fact that our world is particularly parsimonious in this way. Its theory of many-particle systems happens to be strikingly simple, comprising the dynamics of individual particles, a natural probability assumption, and nothing more – no further structure.

Unlike the traditional principle of indifference, the symmetry considerations in play here are not epistemic, a priori, or arbitrary. For the dynamics determines the parameters over which to distribute the probabilities uniformly. The justification is ultimately empirical. It comes from empirical evidence of actual frequencies, combined with the (typically successful) inference that we do not infer more structure than is indicated by the fundamental dynamics. Of course, uniformity over this structure does not *force* the uniform distribution; rather, its success, combined with its simplicity given the dynamical laws, gives us reason for it. For that reason, it escapes the usual problems with relying on indifference to infer physical probabilities.[15]

And once we have the statistical mechanical distribution in place, I claim that it can explain the success of our everyday macroscopic inferences from symmetries to probabilities. These inferences are successful because we live in a world of which statistical mechanics is true.

To see why this seems plausible, consider a simple case like a coin toss. We infer, even before we toss the coin, that each of its physically symmetric sides is equally likely to come up. Repeated tosses of the coin confirm this prediction. Why does our initial inference succeed?

For simplicity, imagine that I am holding the coin balanced vertically on a table. The "toss" consists in my letting go of the coin and its falling to the left (to land heads) or to the right (tails). Think of the phase space of the coin, and consider the region corresponding to its initial macrostate. This region contains all the microstates compatible with the coin's being in this location, with its having this particular size and average temperature,

[15]It also escapes difficulties faced by other approaches, such as Jaynes' (Jaynes, 1983) and approaches based on ergodic theory, neither of which I discuss here.

and so on. Now place a uniform probability distribution over this region, as statistical mechanics tells us to do.

Out of all the possible initial microstates of the coin, think of the different combinations of positions and momenta of the particles that will result in either a heads or a tails outcome. Think in particular of the different possible combinations of momenta: in this idealized case, slight differences in the momentum of even a single particle will determine that the coin falls to the left or to the right.

Statistical mechanics says that, compatible with any given macrostate, such as the initial macrostate of the coin, there are just as many microstates in which a given particle is heading to the left as to the right. (There is a one-one mapping between each microstate and its time-reverse, the same microstate with reversed particle velocities; and for any microstate that realizes a given macrostate, so will its time-reverse.) Since the initial velocities of the particles determine how the coin will land, this means there are just as many ways the initial velocities could be arranged so as to wind up tilting the coin to the left as to the right when I let go of it. In other words, half the initial phase space region will be taken up by microstates such that, if the coin were in one of those, it will fall to the left; half to the right. The uniform distribution with respect to this measure says that any such "left-directed" microstate is equally probable as any "right-directed" one.

According to statistical mechanics, there are just as many initial microstates the coin could be in such that, when I let go of it, it will fall to the left as to the right. This distribution thus counts these microscopic differences in such a way that they add up to an equal probability for each of the two macroscopic outcomes. (The initial distribution, combined with the deterministic dynamics, will yield a similarly uniform distribution over microstates at all times.[16]) Within a bunch of similar coin tosses, statistical mechanics says the microstates will be distributed with approximately half the tosses starting out in "left-directed" microstates, half in "right-directed" ones. At the macroscopic level, this yields the prediction that half the tosses will land heads and half tails; or similarly, that a given toss has a $\frac{1}{2}$ chance

[16]That the initial distribution will remain uniform is a common assumption in statistical mechanics. See Lebowitz (1993), Albert (2000). Note that this should be true for ordinary statistical mechanics as well as Albert's unorthodox version, which conditionalizes these probabilities on the low entropy past.

of landing heads.

That is precisely our initial inference! We infer, on the basis of our macroscopic information, before observing any actual tosses, that the coin has a $\frac{1}{2}$ chance of landing heads. We similarly infer that in a long sequence of tosses, we will get heads about half the time.

This suggests that the *reason* for the success of our initial inference is the truth of the statistical mechanical distribution. Our inferences from symmetries to probabilities succeed when the symmetries we observe happen to match the symmetries in the statistical mechanical probabilities. For when there is this correspondence between the observed symmetries and those in the distribution of fundamental states, the uniform distribution over microstates will yield a similarly uniform distribution over the macroscopic possibilities. It is not that we have know what the statistical mechanical probabilities are; our inferences succeed even though we do not generally know about statistical mechanics. Rather, these inferences turn out to be successful because (and when) the observed symmetries align with the symmetries in the distribution of canonical coordinates. The reason we tend to make these inferences is our past experience, and our past experience has been in a statistical mechanical world.

The above example is admittedly quite idealized. How the coin lands will depend on other factors, such as the velocities of the surrounding air molecules, the angular velocities of the particles in my hand, and more besides. It will take more to argue that the above idea should work even when we include these real-life complications. Yet I think we can reasonably consider it a plausible hypothesis, given the empirical success of statistical mechanics. (Note, though, that this is where Strevens' view might seem more plausible. Strevens (1998) similarly argues that underlying mechanical considerations help explain the success of our symmetry-based inferences to probabilities. Yet his account requires a relatively smooth distribution, not a completely uniform one over canonical coordinates. The above proposal is more ambitious and correspondingly more prone to failure. If successful, however, it could provide a deeper, more unified approach to objective probabilities in general, by explaining the success of many different kinds of probabilistic inferences we make about the world, including the success of Strevens' distribution.)

On this view, the statistical mechanical distribution at once explains the success of our everyday inferences from symmetries to probabilities, and justifies our symmetric probability assignments. We don't rely on some a priori principle to (successfully) infer the frequencies with which actual outcomes occur. Instead, we have learned from experience how systems' microstates are in fact distributed, and we have updated our degrees of belief in what the initial chances are. Any seemingly a priori expectation we might have that the frequencies will match the symmetries in a situation really stems from the experience we have had in a statistical mechanical world.

5. Subjective probability

Nothing so far rules out relying on indifference for determining credences. Given the particular problems we ran into for physical chances, subjective probabilities might seem just the thing for which symmetry considerations could suffice. This seems to be the general assumption: even if indifference can't tell us the objective chances of outcomes, it can dictate what degrees of belief we should have in different possibilities' obtaining.

Can we rely on a principle of indifference here? I will to start to answer this question by looking at a particular view in epistemology called Uniqueness.[17] This will lead to a more general conclusion.

Uniqueness is the view that, given a total body of evidence, there is a unique set of beliefs a rational person can have. Any other belief – any other *degree* of belief – would be irrational.

Though this might seem an implausibly strong view, the idea behind it is intuitive. If there were more than one rationally permissible conclusion, given one's total evidence, then any belief one winds up with must really be *ir*rational. If the total evidence does not uniquely determine a conclusion, then some other, non-evidential factor must have played a role in one's belief formation. But a belief based on such arbitrary factors can't be rational. For such a belief is no more likely to be *true*.

I want to start with this view because it seems committed to the kind of symmetry principle I've been discussing. Here's why. In order for there to be only one rational conclusion given the evidence, as Uniqueness maintains, there must be a uniquely rational set of beliefs one can have before

[17] See White (2005).

getting that evidence. Otherwise, equally rational people with the same evidence could come to hold different beliefs, by updating in the right way[18] from their differing prior beliefs.

To avoid this, Uniqueness must hold that there is a uniquely rational set of priors anyone can have. This means the view needs some basis or principle on which to conclude that certain priors *are* the uniquely rational ones. Since these are beliefs we have before obtaining evidence, it seems this will have to be some kind of a priori principle.

The natural idea is to employ an indifference principle. Before obtaining any evidence, figure out what the different possibilities are, and assign an equal subjective probability to each one. Why equal probability? Without any evidence favoring one possibility over any other, it would be arbitrary to place more weight on any one possibility than any other. The only reasonable thing to do in the absence of evidence is to divide up the possibilities equally. This captures the motivating intuition behind the view: a rational person needs evidence in order to favor one possibility over another; otherwise any such preference would be completely arbitrary. And arbitrary factors cannot be grounds for rational belief.

At first glance, this reliance on indifference seems to avoid the pitfalls we ran into for physical probabilities. Since we are not trying to determine the objective chances of events, we don't run the risk of getting probabilities that fail to match the empirical facts. Here, we are simply using indifference to tell us which initial credences are (uniquely) rational. This seems to be the reason people think indifference should work for credences, if not for chances: whereas empirical frequencies can tell us that the actual chances are different from what we had thought, they cannot tell us that our initial credences were irrational.

This does not, however, manage to escape the familiar difficulties we've seen. First, there remains the problem that any probability assignment will depend on the description of the possibilities. For physical probabilities, the problem was that there is no a priori reason to conclude that any particular parameter space is the uniquely correct one. Here, the problem is that there is no reason to conclude that a particular description is the uniquely *rational* one. But then there can be no unique, non-arbitrary way of assigning each

[18] See note 11.

possibility equal subjective probability, as Uniqueness requires.

Second, even if there were uniquely rational parameters for describing the possibilities, why must we distribute our subjective probabilities uniformly over them? That we are trying to determine our *priors* introduces a new difficulty we didn't face for chances, a difficulty which stems from the reason people think we can use indifference for determining credences. The problem is that we can't rely on empirical evidence to tell us whether these are reasonable degrees of belief about the chances. For we are assigning these probabilities *before* getting any such evidence. But since there is no a priori reason to suppose that an initially uniform assignment is any more likely to lead us to the *truth*, Uniqueness is left without the justification it claims for its (uniquely rational) priors.

It might appear arbitrary, absent any evidence, to put more confidence in one possibility over any other. But it can seem seem just as arbitrary to choose an initially uniform distribution, especially if we cannot show that it is any more likely to yield true beliefs. There may very well be some rational constraints on our priors. Yet it is hard to see why *symmetry* must be one of them – at least, not until we have evidence that this is more likely to yield true beliefs, in which case we are no longer talking about a principle for determining what beliefs we should have before getting any evidence. Of course, symmetries may *affect* our initial credences, and in that sense be a reason for them; nonetheless, prior to getting any evidence, it is no less reasonable to choose an initially asymmetric assignment over a given set of parameters. Either one serves just as well as an initial distribution over credences.

There are examples which bring out our intuition that there are a priori symmetry constraints, at least for certain types of inference. Consider an example of Roger White's.[19] What should we infer is the probability that the number of electrons in our universe is an exact multiple of 10,100? Intuitively, it ought to be very low. Imagine all the possible numbers of electrons there could be, and it seems extremely unlikely that the actual number would be a multiple of 10,100.

Even this inference, however, is presupposing things we cannot know a priori: that electrons in our world do not come in multiples of 10,1000; that

[19] White (ms).

any particular number of electrons is just as likely as any other, regardless of how the world actually is; and so on. Consider a similar inference, familiar from the literature on indifference principles: that a needle is extremely unlikely to land on the floor at any particular angle with respect to the horizontal. Though intuitive and seemingly a priori, this inference, too, is based on our experience in a world like ours, namely, a world with no force field or intrinsic spatial asymmetry picking out a preferred spatial direction or location.

It is these empirical facts which ground our intuition that the probability must be very low. We know from experience that there are these spatial symmetries, just as we know from experience that (as far as we can tell) electrons can exist in any number. I submit that, for any example which seems to demonstrate that symmetry is an a priori constraint on reasonable beliefs about the chances, we are really smuggling in some such empirical assumptions.

Uniqueness in particular needs the link to truth – from symmetrically distributed initial credences to the likelihood of getting true beliefs – in order to ground its assumption that there *are* uniquely rational priors. Without being able to say that these priors are more likely to lead to true beliefs, the view loses its force against more permissive epistemologies. The intuitive pull to Uniqueness is the idea that if the evidence does not uniquely determine a rational conclusion, then one's belief must depend to some extent on arbitrary factors and so be irrational. The belief would be irrational because there is no reason to think that such irrelevant or arbitrary factors are truth-conducive.

We have already seen, however, that uniformity in initial probabilities *is* no more likely to lead to truth. An initially symmetric distribution is a priori no more likely to match the actual facts than some other distribution. Moreover, indifference over the apparent symmetries can yield the *wrong* results. Consider the probability distributions over statespaces of quantum mechanical particles. Whether or not two configurations related by an exchange of identical particles count as the same state depends on the type of particle involved. This demonstrates that we need some evidence to be justified in concluding that a given uniform distribution, over a particular way of counting the possibilities, is correct, even if it appears to us to be uniquely given by symmetries.

Thus, the reason it initially seems that indifference should work for subjective probabilities – that the probability assignment is not immediately answerable to the empirical facts – is the very reason it *can't* work here. We could instead take indifference to be a brute assumption, but this too relinquishes any justification for the constraint via its likelihood of yielding true beliefs. Nor will the consistency or coherence considerations we use to show that our priors should conform to the probability axioms do the job.

The general conclusion is this. Initially, it might seem as though someone with no evidence to distinguish among different possible outcomes, with no reason to to expect one rather than another to occur, ought to infer that each one is equally likely (or almost equally likely), on pain of irrationality. We now see, however, that any view which claims indifference to be a rational constraint on priors runs into the following trouble. Insofar as we link the rationality of a belief to the likelihood of its being true, we cannot rely on indifference: a symmetric probability distribution is a priori no more likely to be truth-conducive than any other. And without this link between rationality and truth, it is hard to see why we should care about being rational in the first place.[20]

6. Conclusion

I conclude that a priori or epistemic symmetry considerations do not suffice as a basis for assigning probabilities. A priori symmetries cannot suffice to tell us what the actual chances of outcomes are; for they cannot tell us what the fundamental possibilities are, let alone that any particular probability distribution over them will be empirically successful. Nor must symmetries factor into reasonable initial credences about the chances of different possible outcomes.

The impression that we can rely on a priori symmetries, stemming from our epistemic situation, really comes from our past experience in a world of which statistical mechanics is true. The symmetries in our world's fundamental dynamical structure, and the correspondence between those funda-

[20]The problem arises for any view that relies on a priori symmetries for rational credences; to the extent that a view allows in other factors, in particular empirical ones, it can escape these difficulties.

mental symmetries and the macroscopic symmetries we observe, explains the success of our everyday inferences from symmetries to probabilities. Absent any such evidence, for truly prior credences, we can go ahead and rely on symmetry to choose a particular initial distribution; but so too can we choose some other asymmetric distribution over initial credences. Neither one is more reasonable than the other, not until we have some experience in the world. For us, not until we have experience in a statistical mechanical world.

References

Albert, David Z. (2000). *Time and Chance*. Cambridge, MA: Harvard University Press.

Jaynes, E. T. (1983). *Papers on Probability, Statistics, and Statistical Physics*. Dordrecht, Holland: D. Reidel.

Landau, L. D. and Lifshitz, E. M. (1980). *Statistical Physics*. Oxford: Butterworth-Heinemann, 3rd edn.

Lebowitz, J. L. (1993). "Microscopic Reversibility and Macroscopic Behavior: Physical Explanations and Mathematical Derivations." *Physics Today* 46, 1–20.

North, Jill (ms). "The 'Structure' of Physics: A Case Study." .

Singer, Stephanie Frank (2001). *Symmetry in Mechanics: A Gentle, Modern Introduction*. Boston: Burkhauser.

Sklar, Lawrence (1993). *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.

Strevens, Michael (1998). "Inferring Probabilities from Symmetries." *Noûs* 32, 231–246.

Tolman, R. C. (1979). *The Principles of Statistical Mechanics*. New York: Dover. Originally printed in 1938 by Oxford University Press.

van Fraassen, Bas C. (1989). *Laws and Symmetry*. Oxford: Clarendon.

White, Roger (2005). "Epistemic Permissiveness." *Philosophical Perspectives* 19(1), 445–459.

—— (ms). "New Thoughts on the Old Problem of Induction." *forthcoming*
.