

# An Empirical Approach to Symmetry and Probability

*Jill North*

*(Studies in History and Philosophy of Modern Physics 41 (2010): 27-40)*

## Abstract

We often rely on symmetries to infer outcomes' probabilities, as when we infer that each side of a fair coin is equally likely to come up on a given toss. Why are these inferences successful? I argue against answering this question with an a priori indifference principle. Reasons to reject such a principle are familiar, yet instructive. They point to a new, empirical explanation for the success of our probabilistic predictions. This has implications for indifference reasoning generally. I argue that a priori symmetries need never constrain our probability attributions, even for initial credences.

## 1. Introduction

I have a die I'm 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, on any one of its six faces. What probabilities should I assign to these different possible outcomes?

It seems reasonable to assign the same probability to each possibility, concluding that each face has a  $\frac{1}{6}$  probability of coming up. This seems like the right physical probability, or objective likelihood, of getting any particular face when I throw the die, as well the subjective probability, or degree of belief, I should have in each one. (You might think that there are no objective probabilities here. You might think that there can be no chances other than 0 or 1 if the laws are deterministic; or that there are no objective probabilities at all in the world. I disagree, but leave this aside

for now.<sup>1</sup> If you have such a view, you can translate my talk of probability into your preferred subjectivist terms.) Since I have no information that the die is loaded, and there appear to be no other asymmetries that would make a difference to the outcome, if I said that the face with number 5 on it has a probability greater than  $\frac{1}{6}$ , this would strike you as an arbitrary, unreasonable preference; I would be irrational to bet at those odds. To prefer one face without any relevant reason is to rely on factors that have nothing to do with the actual state or behavior of the die.

Instead, I should infer that each face has a  $\frac{1}{6}$  probability of coming up. And my degrees of belief should follow suit. The reason to infer symmetric probabilities is the symmetry in the die and my epistemic state with respect to the different possible outcomes.

In cases like these, we reason as follows. First, look at the symmetries in the set-up. Then use these to determine the outcome space, the set of elementary possibilities to which we initially assign probabilities. (The outcome space for the die is  $\{1, 2, 3, 4, 5, 6\}$ .<sup>2</sup>) Finally, assign an equal probability to each such possibility. Distribute probabilities, both objective and subjective, uniformly over the possibility space.<sup>3</sup> Calculate the probabilities of any non-elementary events from these basic probabilities (using the standard probability calculus, say).

We rely on this principle: assign an equal probability to each basic possibility, where the possibilities are given by symmetries relative to our knowledge of the situation. This is an “indifference principle”: infer a distribution that is indifferent among the outcomes which, for all we know, could obtain. This principle underlies our everyday reasoning for the outcomes of die throws, coin tosses, and the like. It *works*. Symmetric probabilities approximate the observed frequencies. A fair die does come up number 5 approximately  $\frac{1}{6}$ th of the time, an unbiased coin lands heads in about half the tosses.

---

<sup>1</sup>Loewer (2001, 2004) argues that a best-system account of laws, for one, allows for deterministic chance. (See Frigg (2008a, forthcoming); Winsberg (2008) for disagreement.) More on setting this aside, below.

<sup>2</sup>We could specify the outcome space in terms of propositions rather than sets.

<sup>3</sup>Strevens (1998) calls this a “non-enumerative statistical induction.” We infer the probabilities by inspecting the symmetries in the set-up, not by tallying up the frequencies with which we observe the different outcomes to occur.

Remarkable. Very little information is needed to infer the right probabilities. Just by inspecting the symmetries in the set-up, before witnessing any outcomes, we make extremely successful probabilistic predictions. This arouses suspicion. Do symmetries alone indicate these probabilities? If not, then why are we so successful in our probabilistic predictions?

One distinguished tradition holds that symmetries do tell us the probabilities.<sup>4</sup> On this view, we can have a priori knowledge of outcomes' probabilities, given the relevant symmetries in the set-up. A similar idea lies in epistemological views claiming that symmetries dictate what initial credences one ought to have. Both these views say that symmetries dictate our initial probability attributions.

I think these views are misguided. Symmetries alone can't explain the success of our probability assignments; nor, I will argue, do they suffice to justify the probabilistic inferences we make. Reasons to reject indifference principles are familiar, but instructive. They suggest a new explanation, at its core empirical. I argue that these inferences are successful because, and when, the observed symmetries align with underlying symmetries in the world, indicated by the fundamental dynamical laws. This tells against any use of indifference, even for constraining our initial credences.

## 2. Symmetric probability: classical physics

Do symmetries explain the probabilities we use in physics, as they seem to do for the outcomes of die throws? An important example suggests so: classical statistical mechanics.<sup>5</sup> Many books and papers on statistical mechanics cite an indifference principle<sup>6</sup> to justify and explain these

---

<sup>4</sup>See Van Fraassen (1989, ch. 12) on this tradition, which traces back to Laplace.

<sup>5</sup>I think that similar conclusions hold for the probabilities of quantum statistical mechanics, but this raises additional technical issues not central here. (See Wallace (2002) for argument that things will be different in that case.) I limit discussion to classical theories. Emch (2007) is a review of issues in foundations of quantum statistical mechanics.

<sup>6</sup>Another approach tries to derive the probabilities from the dynamics using results from ergodic theory. For discussion and references, see Ehrenfest & Ehrenfest (2002); Sklar (1993); Uffink (2004, 2007). The results have not been shown to hold of ordinary systems (Earman & Rédei, 1996); indeed, Goldstein (2001) argues that the statistical mechanical conclusions hold regardless of ergodicity. Nor can the ergodic approach

probabilities<sup>7</sup> (if any underlying principle is offered<sup>8</sup>). Though I agree that symmetries have a role, I disagree on the kind of symmetries at work. That disagreement is important. Once we better understand these symmetries, we can explain the use of these probabilities—in a different, ultimately empirical, way.

First, a refresher on some features of statistical mechanics that we'll need here. There remains lots of disagreement on the proper understanding and formulation of the theory. I cannot do justice to the different approaches here. I will generally assume a Boltzmannian approach, though this shouldn't affect my main conclusions. (This is because the different approaches all rely on some version of the “statistical postulate” that is my focus here.)<sup>9</sup> At worst, take my conclusions as applying to a Boltzmannian framework. At best, take them to apply more generally, while at the same time furthering the cause of that framework.

The fundamental state of a classical system is specified by the positions and momenta of its particles (along with the particles' intrinsic features, like mass and charge). The different possible states of a system

---

avoid the need for some initial probability assumption. Sklar (1973, 1993, 182-188); Leeds (1989); Earman & Rédei (1996); van Lith (2001) discuss the problem and proposed solutions. Malament & Zabell (1980) and Vranas (1998) are improvements on the traditional approach. For more on the variety of approaches to understanding probability in statistical mechanics, both historical and current, see von Plato (1994); Guttman (1999).

<sup>7</sup>As, for example, in Tolman (1979, 59-61); Landau & Lifshitz (1980, 5); Feynman et al. (2006, ch. 6). Tolman takes indifference as a *prima facie* constraint; I argue against this more limited use of indifference, too.

<sup>8</sup>Feynman (1998, ch. 1) posits the probabilities more as fundamental law.

<sup>9</sup>On the idea that the views all rely on a version of this postulate, see Callender (2008). On the “hodgepodge of approaches, formulations, and schools” (Sklar, 1993, 5), see O. Penrose (1979); Sklar (1993); Uffink (1996b); Frigg (2008b). On the development and foundations of statistical mechanics, see Ehrenfest & Ehrenfest (2002); Sklar (1973, 1993, 2000, 2001, 2007); Hagar (2005); Uffink (2007); Frigg (2008a,b). On Gibbs' approach, see Gibbs (1902); Ehrenfest & Ehrenfest (2002); Sklar (1993); Lavis (2005, 2008); Earman (2006); Pitowsky (2006); Uffink (2007). Arguments for Boltzmann's approach are in Lebowitz (1993a,b,c, 1999a,b); Bricmont (1995); Maudlin (1995); Callender (1999); Albert (2000); Goldstein (2001); Goldstein & Lebowitz (2004); against the approach, see Earman (2006). On ergodicity, see note 6. Some textbooks, of varying approaches: Khinchin (1949); Prigogine (1961); Tolman (1979); Landau & Lifshitz (1980); Pathria (1996); Penrose (2005).

are represented mathematically in a statespace. In classical statistical mechanics, the statespace is called *phase space*. For a system of  $n$  particles, the statespace has  $6n$  dimensions, one per position and momentum coordinate per particle (assuming the particles move freely in three spatial dimensions).<sup>10</sup> Each point in phase space represents a different possible exact state, or *microstate*, of the system. A *macrostate*, given by the system's macroscopic features, like the average temperature, pressure, and volume, corresponds to a region in phase space, each point of which represents a microstate that realizes the macrostate.<sup>11</sup> The history of a system is represented by a trajectory through phase space; this is picked out by the dynamics, given the initial state and energy function.<sup>12</sup>

To make predictions about a system, we place a probability distribution over the region in phase space corresponding to its macrostate—the region containing the points representing microstates compatible with the system's macroscopic features. The standard distribution that's used is uniform, with respect to the standard measure (the standard Lebesgue measure, defined over the canonical coordinates).<sup>13</sup> This distribution assigns equal probabilities to equal phase space volumes, on the standard measure, with the result that the probability of a phase space region, such as the region compatible with a system's macrostate, is proportional to its standard volume. This distribution assigns an equal probability

---

<sup>10</sup>In general, the phase space has dimension  $2nr$ , where  $n$  is the number of particles and  $r$  is the number of degrees of freedom.

<sup>11</sup>The set of macrostate regions partitions the available phase space. I leave aside the question of which features specify a system's macrostate. We can assume here that there is a set of macroscopic features, in terms of which there are empirical regularities (such as those of thermodynamics) that we try to explain by means of statistical mechanics.

<sup>12</sup>I assume throughout that a given system is energetically isolated. This assumption is not uncontroversial. It is rejected in particular by advocates of interventionist approaches to explaining thermodynamics on the basis of statistical mechanics.

<sup>13</sup>More on canonical coordinates below. This measure is also called the Liouville or microcanonical measure. Lebesgue measure is the natural extension of volume to high-dimensional spaces. Following Lebowitz (1999b), let  $\Gamma$  be the phase space,  $M$  the system's macrostate, and  $\Gamma_M$  the sub-region of  $\Gamma$  corresponding to  $M$ . The measure of the set of microstates corresponding to a subset of  $\Gamma_M$  is then given by the  $6n$ -dimensional Liouville volume of the subset, normalized by the volume of  $\Gamma_M$ , labeled  $|\Gamma_M|$ :  $|\Gamma_M| = \int_{\Gamma_M} \prod_{i=1}^N d\mathbf{r}_i d\mathbf{p}_i$ , with  $\mathbf{r}_i$  and  $\mathbf{p}_i$  the position and momentum of the  $i$ th particle.

(density<sup>14</sup>) to each microstate compatible with the system's macrostate.

A few caveats. First, there are well-known problems with placing uniform probabilities over the microstates compatible with a system's current macrostate. These problems have to do with explaining the time-asymmetry of thermodynamic behavior.<sup>15</sup> Some try to solve the problem by taking the uniform distribution only over microstates within a proper subset of the region representing a current macrostate. In section 5, I say why one's view on this should not affect things here; in particular, why it shouldn't matter for our purposes whether we assume a current uniform distribution, or one that is uniform at the initial time and conditionalized on past low entropy for all other times. Set this aside for now.

Second, notice that we do not place the uniform distribution over a system's entire, unbounded, infinite-volume phase space, but over the (bounded) region in phase space corresponding to its macrostate.<sup>16</sup> Typically, the constant energy is one such constraint. So that the distribution is uniform over the sub-region of the  $6n - 1$ -dimensional energy hypersurface compatible with the system's other macroscopic features. That is, we take the induced or conditionalized measure on the relevant region of the  $6n - 1$ -dimensional energy hypersurface (the microcanonical measure); alternatively, of the thickened energy shell.<sup>17</sup>

This raises a question. I have been talking as though not only the initial distribution over a phase space region, but one that is restricted

---

<sup>14</sup>For convenience, I talk of probabilities of microstates, represented by points in phase space, where these are really probability densities; the probabilities of microstates are really the probabilities of regions of microstates.

<sup>15</sup>In addition, it is incoherent to apply the uniform distribution at more than one time in a system's history, at least given an objective understanding of the probabilities: see Sklar (1993, 265, 294); Albert (2000, 79-80). Leeds (2003) avoids this problem with an alternative understanding of the probabilities.

<sup>16</sup>The assumption of a bounded phase space is made throughout statistical mechanics. It is crucial to many key results, like Liouville's theorem. This is a reasonable assumption to make for classical isolated systems, like boxes of gas, which have constant energy and are of finite size. I set aside the large question of how, and whether, we can apply this to systems in which gravity is the dominant force (where the phase space may be unbounded): see Earman (1981, 2006); Callender (2008, forthcoming) for discussion.

<sup>17</sup>The thickened energy shell contains microstates for which the system's Hamiltonian lies between energy  $E$  and  $E + \delta E$ . (Thanks to an anonymous referee for pointing this out.)

to a given sub-region (such as a region of the energy hypersurface), will be uniform throughout that sub-region. Why think that an initially uniform distribution will, when conditionalized on a set of macroscopic features, induce a similarly uniform distribution? This is a big question. However, it is commonly assumed that an initial uniform distribution, restricted to such a (possibly lower-dimensional) sub-region, will be at least approximately uniform over that sub-region. I return to this in section 5. For the purpose of this paper, I assume that this is plausible; proofs of such claims in statistical mechanics are notoriously difficult to come by. Yet the dynamics plausibly make the assumption reasonable. Thus, Lebowitz says that, although it “would of course be nice to prove this in all cases...Our mathematical abilities are, however, equal to this task only in very simple situations....These results should, however, be enough to convince a ‘reasonable’ person” (1999a, 521).<sup>18</sup> In any case, and

---

<sup>18</sup>Callender writes of this assumption, needed especially by the Boltzmannian who conditionalizes the uniform distribution on a low entropy past state of the universe:

[T]he Boltzmannian must cross one’s fingers and hope that the dynamics is kind. Recall that any sub-system corresponds, in phase space, to a lower-dimensional subspace of the original phase space. The hope must be that when we project the original approximately uniform distribution onto this subspace and renormalize we again find a distribution that is approximately uniform. Pictorially, imagine a plane and a thoroughly fibrillated set of points on this plane, a set so fibrillated that it corresponds to an approximately uniform measure. Now draw a line at random through this plane and color in the points on the line that intersect the fibrillated set. Are these colored points themselves approximately uniformly distributed? That is what the Boltzmannian needs for local thermodynamic systems, except with vastly many higher dimensions originally and much greater dimensional gaps between the phase space and subspaces.

How are we to evaluate this reply? Based on experience with many types of systems, some physicists don’t balk at the thought of such fibrillation. They see it in some of the systems they deal with and in the absence of constraints ask why things shouldn’t be so fibrillated. The unhappy truth, however, is that we simply have no idea how to evaluate this claim in general (2008, 25).

Even so, Callender says, “My own attitude is to note this assumption as a large one and move on.” This is what I will do here. (Note that this is where ergodic results can be useful, by helping to make this assumption plausible.)

A referee objects that the standard restricted measure isn’t uniform over the relevant

most important for our purposes here, the initial distribution, taken over a given macrostate phase space region, is generally assumed to be uniform, regardless of whether or not a relevantly conditionalized distribution will be likewise uniform throughout a sub-region. And it is this initial uniformity that is my focus here. Set this, too, aside for now.

There is no question that these probabilities are empirically successful.<sup>19</sup> Consider the statistical-mechanical explanations of thermodynamic

---

sub-space, for it is divided by the gradient of the Hamiltonian; cf. Kac (1959, 63), Frigg (2008b, 180, eq. 3.46). (Though Jos Uffink points out that the microcanonical distribution on the thickened energy shell will be uniform over the canonical coordinates of the shell.) Yet as far as I can tell, the restricted distribution is often assumed to be approximately uniform over a sub-region (for typical macro systems: note 22), while it is also acknowledged that we don't have a proof. Sklar (1993, 95-96) writes that, assuming the distribution is absolutely continuous with the standard measure, "a single probability function can be obtained that 'spreads the probability' over the appropriate region of phase-space so that the total probability assigned to a region is obtained by a measure of the amount of the total probability over the whole phase-space that is spread in the region in question. Most commonly in statistical mechanics it is 'uniform' spreading that is posited, the probability being assigned to a region of micro-states just being proportional to the size of that region in the chosen measure." To put it another way, it is commonly assumed that an initial set of phase points will at later times be "fibrillated" over the available phase space (Sklar, 1993, 233-234). In any case, even if the restricted distribution isn't uniform, it is typically obtained by restricting an initial distribution that is. Similarly, according to the past low entropy approach, uniform probabilities hold only at the beginning of the universe, at other times approximating uniform ones.

<sup>19</sup>Although there *is* a big question about the range of phenomena they are successful for. Jos Uffink objects that the above recipe works only for systems in thermal equilibrium. For we can prepare systems not in equilibrium, like vials of gas released into larger containers, for which predictions using the standard measure (over the canonical coordinates of the larger container) will generally be wrong, at least until equilibrium is reached. This is true when directed at the techniques of equilibrium statistical mechanics. And throughout the discussion that follows, we may assume that the system in question is in local equilibrium. It is generally agreed that the uniform distribution over the system's macrostate region works for equilibrium statistical mechanics; for example, to calculate the values of its macroscopic features at equilibrium. I happen to think that we can extend these probabilities to non-equilibrium systems too, though this is more contentious. For argument that we can so extend them, at least given a Boltzmannian conception of entropy, see Callender (1999); Lebowitz (1999b); Goldstein & Lebowitz (2004). Arguably, the above recipe should succeed for predicting non-equilibrium behaviors, such as that the gas will spread out until it reaches the equilibrium macrostate compatible with the new volume constraint, even if it won't yield the right macroscopic

phenomena. In equilibrium statistical mechanics, these probabilities are used to calculate the values of a system's macroscopic features at thermal equilibrium. The probabilities can also be used to explain the tendency of isolated systems to increase in entropy—of a gas to spread out in its container, of a cup of coffee to cool down. Part of the explanation of this tendency is that, out of all the microstates compatible with a system's current macroscopic constraints, the overwhelming majority, on the standard measure, lie on trajectories that deterministically take the system into higher-entropy future macrostates. A uniform probability distribution with respect to this measure then says that the entropy-increasing microstates are overwhelmingly probable.<sup>20</sup> (Conditionalizing on the energy constraint, the distribution says that the microstates realizing the equilibrium macrostate take up the overwhelming majority of the hypersurface area, and so are overwhelmingly probable.)

But what explains the success of the uniform distribution? Why are these the right probabilities to use? This isn't obvious. Since the set of microstates compatible with a given macrostate is continuously infinite, there are many probability assignments we could use. (I return to this.) Why this one? Similarly, why the standard volume measure? Different measures of phase space regions will likewise result in different probabilities for a system's possible microstates.<sup>21</sup>

---

features of the gas until it reaches equilibrium. For, plausibly, the overwhelming majority of microstates on the energy hypersurface, on the standard measure, will increase in entropy to the equilibrium macrostate and then stay there. Showing that this succeeds generally (and in detail, such as predicting the rate of approach to equilibrium) remains a large project. I take it that recent work (Albert (2000, 2008) being primary examples) has made this plausible (for more discussion see North (forthcoming)). If you disagree, consider the discussion as limited to the standard probability measure of equilibrium statistical mechanics.

<sup>20</sup>This type of argument is relatively standard (though the interpretation will be very different on a Gibbsian conception of the probabilities). Proofs are difficult to come by. I take it that the claim is plausible, which isn't to deny the work in foundations of statistical mechanics aimed at giving it more conclusive status. See also note 19.

<sup>21</sup>Of course, when people question the probabilities of statistical mechanics, they often have more in mind than wondering about uniformity. Some wonder whether these probabilities are unique; I say later why I think that this worry is misguided. Others wonder about the nature of the probabilities, especially given the deterministic dynamics. This is a large issue, which I leave aside; I say below why it shouldn't affect

At this point, you might worry about the success that we are trying to explain. You might worry that the above method for grounding thermodynamic phenomena like entropy increase can't always work. Thermodynamic generalizations like the second law are not strictly true. They hold probabilistically, and only for energetically isolated and large enough systems.<sup>22</sup> Even once a system has reached thermal equilibrium, for instance, there will be fluctuations out of equilibrium. Indeed, given the time reversal invariance and determinism of the underlying micro-dynamical laws, we know that entropy increase will not always hold.

In reply, keep in mind that our goal is to explain these probabilities, and their grounding of various macroscopic behaviors, *insofar as these macroscopic behaviors hold*. This is not to deny that systems fluctuate out of equilibrium; that they can exhibit short-term anti-thermodynamic behavior; that non-isolated systems can reliably decrease in entropy; and so on. I take it that we already know, from thought experiments like Maxwell's demon, and from the reversibility and recurrence objections of Poincaré, Loschmidt, and Zermelo, that the generalizations grounded by statistical mechanics are not strict, but probabilistic; and even then, that they hold only given that we do not have the epistemic access and manipulative capacities of a Maxwellian demon. The question is: why do these probabilities work in grounding macroscopic generalizations, to the extent that those generalizations hold?<sup>23</sup>

Enter indifference. Any system is in one of the microstates compatible

---

my conclusions here. My focus is on why the distribution is reasonable and successful, and to suggest that these same reasons underlie the success of our everyday probability inferences.

<sup>22</sup>This is not uncontroversial; see Callender (2001) for argument. A referee mentions the one-dimensional harmonic oscillator. Its motion is periodic and won't be accurately predicted by the uniform distribution; some phase points are preferred (Penrose, 2005, 41). I take it that the goal is to explain the probabilities' grounding of macroscopic behavior. And we know that some systems' motions, like that of the one-dimensional harmonic oscillator, won't fall under the thermodynamic generalizations that we use statistical mechanics to explain; think of a few isolated, non-interacting particles bouncing back and forth in a box. The intuition is that such systems are rare or special, the more so the larger the number of particles. As Sklar (1993, 241) puts it, "plainly, the large number of micro-components is essential for the thermodynamic behavior of systems."

<sup>23</sup>More contentiously, statistical mechanics may be able to predict the anti-thermodynamic behavior as well, as Albert (2000, 2008) argues.

with its macrostate. And we have no reason to think it is more likely to be in one rather than any other. Conclusion? Assign an equal probability to each possible one.<sup>24</sup> This is akin to our reasoning for the die. Absent reason to the contrary, weight each possibility consistent with our knowledge equally. (Hence the ‘hypothesis of equal a priori probabilities’ in physics books.) It seems we can figure out what probabilities to use in statistical mechanics from a priori indifference reasoning, just as we seem able to do in our everyday inferences about die throws. Initial symmetries seem to justify our probability attributions and to explain their success.

There are two general reasons that this can’t be right. The first is familiar, but worth mentioning in order to motivate an alternative. Probability assignments depend on how the possibilities are described. And in general there is no a priori way of picking out a unique and non-arbitrary set of parameters to use.<sup>25</sup> We must choose a parameterization in order to assign probabilities, and indifference is no help here. It says, *given* a parameterization, distribute probabilities uniformly. It does not also say which parametrization to use; and there is no a priori reason to think that any one is correct. Indifference, then, just won’t say what the probabilities are. At best, it gives probabilities relative to an arbitrary choice of description.<sup>26</sup>

---

<sup>24</sup>The standard volume measure also seems to give equal sizes to macrostates with the same “number” of microstates. I return to this in section 4, focusing here on uniformity.

<sup>25</sup>An example (van Fraassen, 1989, 303): A factory makes cubes with side length  $\leq 1$  foot. What is the probability that a cube’s side length is  $\leq \frac{1}{2}$  foot? The answer depends on whether we distribute probabilities uniformly over side length ( $\frac{1}{2}$ ), area ( $\frac{1}{4}$ ), or volume ( $\frac{1}{8}$ ). Knowing more about the cube factory could do the job, but this just reinforces the conclusion that indifference alone will not. Other familiar examples are Bertrand’s paradox, Buffon’s needle, von Mises’ water and wine case (von Mises, 1981, 66-81). See van Fraassen (1989, ch. 12) for discussion and references. Nor will indifference suffice for the die: compare the probability of its coming up even vs. odd, prime vs. non-prime, one vs. non-one (Sklar, 1993, 199). Sklar notes that if the basic possibilities must be indecomposable events, then indifference will yield unique probabilities; and this differs from the infinite case, where the basic events all get probability zero, on any distribution (note 26). Even then, though, we need an independent sense of indecomposability, and this won’t be given a priori.

<sup>26</sup>Things are worse for infinite possibility spaces like phase space. The basic probabilities won’t determine the probabilities of infinite (measurable) sets. At best, uniformity yields the zero-probability outcomes. Take a countable set of possibilities. Equal finite

There is a second reason that we cannot infer the statistical mechanical probabilities from a priori indifference, a reason that I take to be just as decisive, but which the defender of indifference admittedly needn't buy. We use these probabilities to successfully predict and explain the observed frequencies of outcomes. And it is hard to believe that symmetries could tell us a priori about this kind of empirical matter. Of course, it is too strong to object that symmetries do not give a logical guarantee of the right results. Indifference may simply be an initial, defeasible basis for inferring probabilities, constraining what we ought to initially infer. Still, there seems no a priori reason to think the actual frequencies likely to track the symmetries—no a priori reason that nature will not prefer some microstates to others—so no a priori reason to think we must set our credences by the symmetries.<sup>27</sup> The proponent of indifference may deny this (I return to it in the next section), but should at least grant it intuitively odd to think that symmetries based on what we know of a system rationally constrains, a priori, what to predict about it.

I think that these two reasons derail the use of indifference for assigning probabilities in physics. Even if symmetries could yield non-arbitrary, unambiguous probabilities (answering the first problem), this won't suffice to say what the probabilities are (given the second problem), let alone why our symmetric inferences are successful. Consider a finite, discrete statespace, such as for the six faces of a die. Here there is a unique

---

probabilities violates a probability axiom; zero probabilities violates countable additivity (likewise for infinitesimal probabilities: the sum won't converge to 1); and even dropping countable additivity, there is no unique uniform distribution. For uncountable spaces, think of the points on the real unit interval. Even assuming uniformity, the probabilities of the points won't determine the probabilities of sub-intervals ( $(r_1, r_2)$ ,  $r_1$  and  $r_2$  distinct reals between 0 and 1). For there is no unique measure of the intervals to begin with.

<sup>27</sup>Analogous objections are in Strevens (1998); Albert (2000, ch. 3); Goldstein (2001); Loewer (2001); and in a different way Ismael (2009). See van Fraassen (1989, ch. 12) for more discussion. Similar objections apply to other attempts to salvage indifference. Thus, Jaynes (1983) argues that the distribution must maximize entropy relative to the macroscopic constraints and be invariant under scale changes. But a uniform distribution over one parameterization won't be invariant to all scale changes; and there is no a priori way of saying what the right scale changes are: Milne (1983). (See Uffink (1996a) against the entropy constraint.) Jaynes suggests that it comes from the geometry of the situation, but neither will this be a priori: Marinoff (1994). (See Shackel (2007) against Marinoff's attempt to salvage (limited uses of) indifference.)

uniform distribution; yet even so, systems' microstates may not actually be distributed in this way. Hence we seem no more a priori justified in using indifference-based probabilities rather than others. It is true that symmetric probabilities and the standard measure both seem perfectly *natural*. But this shouldn't mislead us into thinking that they are thereby a priori justified. If you reply that symmetries rationally constrain our credences, providing some initial evidence of the probabilities, then you should at least grant that this is intuitively odd. In section 4, I suggest an alternative.

### 3. Symmetric credence

Even if indifference cannot tell us the (objective) physical probabilities of outcomes, it seems open to say that it constrains rational belief, dictating what (subjective) credences one ought to have. I think that this fails for similar reasons. We can see this by looking at a particular epistemological view relying on indifference. This will lead to a general conclusion about indifference and rationality.

The view is Uniqueness.<sup>28</sup> Given a body of evidence, there is a unique set of (degrees of) belief a rational person can have. Though an awfully strong view, the key idea is intuitive. If there were more than one rationally permissible conclusion, given one's evidence, then any belief one winds up with must really be *irrational*. If the total evidence fails to 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 arbitrary factors cannot be rational. For such a belief is no more likely to be true.

There must also be a uniquely rational set of beliefs one can have *before* getting evidence. Otherwise, equally rational people with the same evidence can come to different beliefs, by updating their differing priors. (Granting a uniquely rational method of updating, assumed here to be something like Bayesian conditionalization.) The view needs a basis on which to say that some priors are uniquely rational. A natural idea is indifference. Uniform probabilities seem to avoid arbitrary preferences.<sup>29</sup>

---

<sup>28</sup>See White (2005a).

<sup>29</sup>It's hard to see what else would do. Conditionalizing on vague priors will yield

This faces the above two problems. The probabilities we assign depend on how the possibilities are described, and there's no a priori reason to think that one description is uniquely rational. If so, then there is no uniquely rational, non-arbitrary way to assign equal subjective probabilities, as Uniqueness requires. And even if there were a unique rational description, why must we distribute credences uniformly? Empirical evidence won't say if some priors are most reasonable, and there's no a priori reason to think that uniform credences are more likely to yield truths about the frequencies.

Now, in saying that it is reasonable to think these priors likely to succeed, the view may just mean "likely according to the correct prior distribution." The problem is that, without measuring a belief's rationality by its likelihood of leading to empirical frequencies, Uniqueness loses its edge over more permissive rivals. The key idea is that if the evidence does not uniquely determine your belief, then it must be based on arbitrary factors and thus irrational—irrational because no more likely to be true. Yet we have seen reasons to doubt that uniformity *is* more likely to yield truth. More, we know that it can give the wrong results. In quantum mechanics, the basic possibilities themselves aren't given by initial symmetries; whether an exchange of identical particles yields a distinct state depends on the type of particle. In other words, which factors are irrelevant or arbitrary is not something that is a priori.

We do have a strong intuition in favor of symmetry constraints. What should we infer is the probability that the number of electrons in our universe is an exact multiple of 10,100?<sup>30</sup> Intuitively, it should be very low. Imagine all the possible numbers of electrons there could be, and it seems extremely unlikely that the actual number is a multiple of 10,100. Yet even this inference is presupposing things that we cannot know a priori: that electrons in our world do not come in multiples of 10,100; that any number of electrons is as likely as any other, regardless of what the rest of the world is like. Take a similar inference, familiar from the indifference literature: a needle is unlikely to land on the floor at any given angle with

---

equally vague posteriors. Norton (2008b) further argues that the state of ignorance described by indifference is incompatible with Bayesian updating, and so is not even representable as a genuine probability distribution. See also Arntzenius (2008).

<sup>30</sup>The example is from White (2005b).

respect to the horizontal. Though intuitive and seemingly a priori, this inference is based on experience in a world like ours, with no force field or spatial asymmetry picking out a preferred direction or location. I submit that any case claiming to demonstrate symmetry as a rational constraint on initial belief really smuggles in some such empirical assumptions.<sup>31</sup>

Here is the general conclusion. Intuitively, anyone with no evidence, with no *reason* to expect one outcome over another, should infer each possibility equally likely, on pain of irrationality. However, any view claiming this as a rational constraint on priors (versions of objective Bayesianism, the bygone Carnapian program, Uniqueness) faces a dilemma.<sup>32</sup> Insofar as we link the rationality of a belief to its likelihood of being true, we needn't rely on indifference: symmetric probabilities are a priori no more likely to be truth-conducive. If we instead say that asymmetric priors are nonetheless irrational, we give up the tie between a belief's rationality and its likelihood of yielding truth; and then it is hard to see why we should care about being rational in the first place. Of course, we need some initial credences to be able to start taking in evidence. The point is not that we are unjustified in choosing any priors at all, but that an initial, pre-evidential, asymmetric set of priors is as good as any symmetric one. In the absence of evidence, there is no *more* reason to use a symmetric distribution rather than some other.<sup>33</sup> So choose one, and update your priors in the right way as the evidence comes in.

#### 4. An empirical approach

I think that we can avoid the unwelcome consequences of a priori symmetry constraints on our probability attributions. Return to statistical mechanics. What tells us to use the uniform distribution here?

Part of the answer is, simply, empirical input from the world: the actual frequencies we use to confirm any probabilities in physics. Using these

---

<sup>31</sup>Likewise for anthropic principles. There is no a priori reason we must use an initial uniform distribution over possible worlds, from which to infer that our world is unlikely.

<sup>32</sup>The problem is for any view claiming that a priori symmetries constrain rational credence. Other initial constraints on credences may escape these difficulties.

<sup>33</sup>Compare von Mises (1981, 75) who says (in arguing against subjectivism) that, "if we know nothing about a thing, we cannot say anything about its probability."

probabilities, we make extremely successful predictions about statistical mechanical phenomena.

This is only part of an answer, since the empirical frequencies underdetermine the exact form of the distribution. Other, relatively smooth, not-completely-uniform distributions should yield just as good predictions for statistical mechanics.<sup>34</sup> Why the uniform one?

I think that we have an answer to this, though we need to back up a bit and work up to it. Recall that the basic possibilities in statistical mechanics are given by the different combinations of particle positions and momenta. The uniform distribution counts each such combination as equally likely. You might wonder why we use these coordinates to describe the possible states to begin with. Aren't we back at the problem of there being no unique, non-arbitrary set of parameters over which to distribute probabilities? Why uniformity with respect to these coordinates and not others?

Set aside statistical mechanics for a moment and think of the underlying particle dynamics. Ask a similar question: why use the position and momentum coordinates to describe the states of the particles?

Here we have an answer. These coordinates yield a simple formulation of the classical dynamical laws, Hamilton's equations. Hamilton's equations are formulated in terms of the generalized position and momentum coordinates, called *canonical coordinates*. Generalized coordinates are any set of independent parameters that completely specify a system's state. Different kinds of generalized coordinates can be used in the very same form of the laws. The generalized "positions" needn't be ordinary positions, for example; they can have dimensions of length squared, or energy, or be a dimensionless quantity. What's required is that the two sets of coordinates be related in the way allowed by, or coded up in, the dynamical equations (as canonical conjugates). The allowable sets of generalized coordinates are the ones that preserve the equations of motion; the ones that are invariant under canonical transformations.<sup>35</sup>

---

<sup>34</sup>See Strevens (1998); Albert (2000); Goldstein (2001); Maudlin (2007); and below.

<sup>35</sup>The symplectic form and Hamiltonian  $H$  together determine a vector field,  $X_H$ , on the phase space.  $X_H$ , in turn, defines a flow on the symplectic manifold, and by Liouville's theorem, Hamiltonian flows preserve the volume form: Hamiltonian flows are canonical transformations. See Abraham & Marsden (1980, 3.3).

Here is another important feature of the particle dynamics. Formulating the equations in this way, with these coordinates, requires a certain mathematical structure, possessed by the theory's statespace. This structure is given by the invariant quantities of the dynamical laws under allowable sets of coordinate transformations, in the same way that the structure of a theory's spacetime is given by invariant quantities under allowable transformations. In Hamiltonian mechanics, this is symplectic structure: the statespace of Hamiltonian mechanics is a symplectic manifold. This is a manifold with a mathematical object defined on it, a symplectic form<sup>36</sup> (the invariant of the equations), which encodes the equations of motion, and a scalar function, the Hamiltonian, which encodes the dynamical features, such as the energy and forces on the system. (The equations can also be formulated without mentioning coordinates; this more directly brings the intrinsic statespace structure to light.)

There are other versions of the dynamics, such as Lagrangian mechanics, which use different sets of coordinates and different statespaces. The Lagrangian formulation even uses two sets of generalized coordinates (the generalized positions and their first time derivatives, the generalized velocities). So we seem to be back at our earlier question: why assume the canonical coordinates of Hamiltonian mechanics?

This is a big question, the answer to which I only outline here.<sup>37</sup> The idea is this. The Hamiltonian formulation, given in terms of the canonical coordinates and a symplectic statespace structure, is, in an important sense, *simpler*. The sense of simplicity that I have in mind is this: the least amount of structure needed to formulate the dynamical laws. Modern formulations of classical dynamics are defined on a statespace, equipped with a certain structure. Different theories' statespaces can differ in that structure. Comparing statespace structures then gives a measure of theories' relative simplicity.

We can compare statespace structures in the same way we compare mathematical structures: by how many "levels" of structure are needed to define a space, starting from a set of points. The levels are ordered according to which mathematical objects must be presupposed to define

---

<sup>36</sup>A symplectic form is a closed, nondegenerate, antisymmetric 2-form.

<sup>37</sup>See North (2009a,b) for detailed discussion of the mathematical structures required by different formulations of the classical dynamics, including the Newtonian one.

others.<sup>38</sup> In this sense, a topological space has more structure than a “bare” set of points; a topology adds a level of structure, indicating the open subsets. A metric space has more structure than a topological space; a metric adds a level of structure, giving distances between nearby points. (A metric induces a topology; a topology doesn’t give metrical relations.) Note that, in general, *symmetries* mean less structure. Given a set of points, another level of structure is needed to pick out a preferred element, giving the space asymmetry. A Euclidean plane with a preferred direction has more structure than a Euclidean plane without one, for example; picking out a preferred direction requires an additional mathematical object, an orientation.

In this sense of comparative structure, the Hamiltonian statespace is simpler than the statespaces of other formulations. Compare the Lagrangian and Hamiltonian statespaces. The Hamiltonian statespace has a volume measure;<sup>39</sup> the Lagrangian statespace has a distance measure. There is a clear sense in which a space with metric structure has *more* structure than one with just a volume element. Metric structure determines or presupposes a volume structure, but not the other way around (in the same way that a metric presupposes a topology, but not the other way around). Intuitively, knowing the distances between points in a space gives you the volumes of regions, but the volumes won’t determine the distances.<sup>40</sup> Metric structure is an *additional* level of structure.

Why is “amount of structure” any measure of simplicity or reason to prefer a theory? I don’t have an argument for this, other than to note that it is something we generally do in physics. We generally infer the least structure needed to formulate the fundamental dynamics—both the mathematical structure used to state the laws, and the physical structure of a world described by those laws. In particular, we infer symmetries in the world from mathematical symmetries, or invariances, in the dynamics. This is clear in the case of spacetime symmetries. From time-translation invariance of the laws, we infer that time has no preferred temporal location. From Lorentz invariance, we infer that spacetime has no absolute

---

<sup>38</sup>Compare Burke (1985, 37); Isham (2003). See Sklar (1974, ch. II).

<sup>39</sup>The symplectic form determines a volume form: Arnol’d (1989, 206). Though note that a symplectic form is different from, and stronger than, a generic volume form.

<sup>40</sup>See Schutz (1980, ch. 4).

simultaneity structure. If the dynamical laws do not require some structure, if they are invariant under transformations that alter the structure, we infer that there is no such structure in the world according to the theory. The inference is not conclusive; we can't be certain that the minimal structure required by the dynamics is correct. Yet physics has done well by this methodological principle. It seems a reasonable guide to when we have inferred the world's structure correctly.<sup>41</sup>

You may balk at using this principle for statespace, even if you agree with its use for spacetime. To the extent that we wish to minimize structure, we want to minimize structure in the world. And a theory's statespace seems instrumental, a formal tool; not something about the world. Minimizing statespace structure then shouldn't matter to the physics.

I disagree. A theory's statespace does tell us about the world. It may do so directly, because the statespace is part of the ontology, as the substantialist says that spacetime is part of the ontology. Or it may do so less directly, telling us about the world in the way that the laws do; for the statespace is the abstract geometric formulation of the laws. Regardless of your view of its metaphysics, a theory's statespace is part of its genuine content. It is an essential component of our best geometric formulations of the physics. Other things equal, we prefer the simplest formulation of a theory. And less statespace structure means a simpler formulation, requiring less mathematical structure and ascribing less structure to the world. (In the same way that less spacetime structure means a simpler formulation, regardless of your views on the metaphysics of spacetime.)

Spot me the inference to the minimal structure required by a theory's dynamical laws. Where does this get us in statistical mechanics?

I claim that this inference lies behind the standard distribution of statistical mechanics. The phase space of statistical mechanics is parameterized by the coordinates we use to describe the states of the particles. Phase space then inherits the mathematical structure of the classical dynamical statespace. This structure, we've seen, includes a volume element: there is a natural volume measure induced by the canonical coordinates.

---

<sup>41</sup> Compare Sklar (1974, 48-49). Earman (1989, 46) considers this a condition of adequacy on dynamical theories. There remains the big question of simplicity considerations in theory choice. To the extent that we think invariances in the dynamics aren't arbitrary but track features of the world, we can sidestep this: see K. S. Friedman (1976).

A probability distribution that's uniform over the statistical mechanical phase space, with respect to the natural volume measure of the dynamics, is then strikingly simple. It requires no mathematical structure beyond what is needed to formulate the dynamics. A non-uniform distribution, or a distribution with respect to a different measure, would require further structure, a kind of preferred-point or -region in phase space structure—just as a Euclidean plane with a preferred direction requires an orientation. A further mathematical object would be needed to pick out the points to be weighted more heavily, when no such preference is given them by the dynamics. Since we do not need any additional structure for the dynamics of the particles in these systems, and since we can formulate a successful statistical mechanics without it, we should infer the uniform distribution as the one that accurately reflects the underlying nature of the world, just as we generally infer that the simplest formulation of the dynamics reflects the world. No more structure is needed to formulate the theory, and no more structure is needed in a world governed by the theory.

This gives us two reasons for the uniform probability distribution, with respect to the standard measure, in statistical mechanics. (1) It yields empirically successful predictions. (2) It is uniform over the structure needed for the dynamics—uniform with respect to the natural volume inherited from the dynamics—and so requires no structure beyond what is “already there” for the dynamics. It is the simplest, most natural, most mildly-constrained distribution we could use, given the dynamics of the particles in these systems.<sup>42</sup> Although we must ultimately justify the distribution by its empirical success, as the case against indifference shows, uniformity over this statespace structure is a further reason for it. From among the candidate distributions yielding correct predictions, the uniform one is simplest, relative to the structure of the dynamics. Although other, relatively smooth distributions should be as successful in

---

<sup>42</sup>The same considerations support the view of Dürr et al. (1992b,a); Maudlin (2007), which requires less structure than a probability distribution. (Volchan (2007) is a different version of the view. Albert (2008) is a reply.) Consider the above as an argument for the uniform distribution, on the assumption we need a probability distribution; if not, these considerations preference their view. The above also supports the proposal of Albert (2000, ch. 7) that if GRW quantum mechanics is true, we'd have an even simpler account, eliminating statistical mechanical probabilities altogether. (Price (2002a,b) disagrees; see North (2002) for a reply.)

statistical mechanics, this is the simplest and most natural we could use.<sup>43</sup>

What if uniform distributions over differently-coordinated statespaces, for different formulations of the laws, disagree on the probabilities of microstates?<sup>44</sup> This needn't worry us here. It is reasonable to suspect, given the above, that the Hamiltonian formulation is simplest, and thus the one we should infer. However, if there do turn out to be different formulations that are equally simple, in the above sense, then it is reasonable to suspect that uniform probabilities over the different statespaces won't yield differences at the level of statistical mechanical predictions, for the reasons that any sufficiently smooth distribution should work for statistical mechanics.<sup>45</sup>

You might object that no more structure need be added to the theory for a non-uniform distribution. You might think that we needn't add structure picking out which points are preferred; we need only say that there are some regions whose probabilities differ from the standard uniform ones.<sup>46</sup> But in order to assign non-uniform probabilities in a way that's empirically adequate, in a way that's empirically confirmable by the evidence we have for ordinary statistical mechanics, we will have to pick out the preferred regions or points, in the same way that we would have to add mathematical structure to pick out a preferred direction in

---

<sup>43</sup>All the more so if we need probabilities in somewhere our physics, as argued by Albert (2008) and Ismael (2009). Note the above is a version of Sklar's (1973, 212) "simple justification," with an added ingredient. Start by guessing at probability distributions, and see whether they make the right predictions. If one does, then we have chosen correctly; if not, we keep guessing until we find the one that does. The additional ingredient here, naturalness relative to the dynamics, selects a distribution from those getting the statistical facts right.

<sup>44</sup>As should be the case for Hamiltonian and Lagrangian mechanics. The Hamiltonian coordinates are related to the Lagrangian ones by a (non-measure preserving) Legendre transform. So a uniform distribution over the Hamiltonian statespace, with respect to the Lebesgue volume element in its coordinates, needn't be uniform over the Lagrangian statespace, with respect to its volume element. It has also been shown that there are alternative equivalent Hamiltonian descriptions, not related by a measure preserving transformation to the canonical ones (Ercolessi et al., 2002); though these, too, seem to presuppose a symplectic structure. I thank Branden Fitelson for discussion and references.

<sup>45</sup>Note 34.

<sup>46</sup>I thank Ted Sider for this question.

a Euclidean plane. A physical theory which says that some microstate-regions are preferred, without also saying which ones—a theory which says there’s a region with twice the probability of another, say, without saying which region this is—won’t yield predictions resembling those confirmed by statistical mechanics as we know it. In order to posit a non-uniform distribution, either we will have to add structure picking out the preferred regions, or end up with a theory whose confirmability by the empirical evidence is questionable. Either way, general scientific grounds tell against doing so.

Remember the two big problems for indifference. There’s no a priori way of picking out a set of parameters with which to characterize the possibilities (and statespace structure); and even if there were, there’s no a priori reason to distribute probabilities uniformly over the possibilities. The proposal here avoids both problems. The empirically confirmed dynamics says which coordinates to use to characterize the possibilities.<sup>47</sup> (Just as empirical evidence says what count as distinct possibilities for quantum mechanical particles.) And a general methodological principle tells us to use a uniform distribution over these possibilities.<sup>48</sup> Note the emphasis on *empirical* considerations. The structure principle comes into play once we have a range of empirically confirmed distributions. If empirical evidence were to disconfirm the uniform one, then simplicity considerations would pick out some other from among the remaining empirically adequate candidates. Note also that the methodological principle is itself justified by a kind of empirical, if inductive, argument: it has generally yielded successful theories for our world.

To be clear, there is no a priori reason that this must be the right probability distribution to use. Nor is it forced on us by the dynamics; I do not claim that it is uniquely natural in being derivable from the

---

<sup>47</sup>Ismael (2009) argues that any such parameterization must be empirically grounded.

<sup>48</sup>This is similar to Strevens (1998, 241): “It seems to be the case that, for whatever reason, our standard variables are 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.” My view differs in proposing that the distribution isn’t just smooth but uniform over dynamical coordinates, and that there’s a reason for this: statistical evidence plus a successful methodological principle. See section 5.

dynamics.<sup>49</sup> I do claim that it is reasonable, even so, given empirical evidence of its success and its naturalness relative to the dynamics. The justification bottoms out in the contingent fact that our world seems to be parsimonious in this way. Its theory of many-particle systems is strikingly simple, including the dynamics of individual particles, a natural probability assumption, and no further structure.<sup>50</sup> If you go on to ask why we should live in such a simple world, I don't have more to say. Our world didn't have to be one for which a natural probability assumption is the one to use for making predictions about its sub-systems. Our world happens to exhibit this simplicity; there is no further explanation of this fact.

Thus, whereas many people think we must prove that this is the only invariant probability measure—the only one that's “carried along” or preserved by the dynamics—on my view this is a feature of the distribution that is chosen for other empirical reasons.<sup>51</sup> It could have been that the probabilities changed with time (say, the probability of microstates for gases' remaining in corners of their containers increased every year). Or it could have been that no useful probability distribution at all was to be

---

<sup>49</sup>In particular, I don't claim to derive the probabilities from the dynamics in the way sought by ergodic approaches (note 6). There the goal is to locate, as Sklar (1973) puts it, more than a reason to think that the distribution gives the right results, but a reason to think that it should give the right results. On my view, one distribution is natural, but not by derivation from the dynamics. See Ismael (2009) for general argument that the probabilities needed in physics are over and above the dynamics. Another worry is that it begs the question to “assume the actual microstate is always in the set of typical points” (Callender, 1999, 371), positing the probabilities to explain them by means of their success. But we justify these probabilities in the same way we do any fundamental physical posit, by accounting for the phenomena in a simple way. We don't need a further reason to eliminate such posits altogether, as argued for in a different way in Callender (2004).

<sup>50</sup>Plus (at least) one more component. In my view, the theory also needs a statement about initial low entropy in order to ground thermodynamic and other macroscopic asymmetries. Which additional component, if any, we need to ground thermodynamics is a large separate question.

<sup>51</sup>The natural volume induced by the canonical coordinates is preserved by the dynamics: Arnol'd (1989, 204-207). Whether a non-invariant distribution could be said to assign genuine probabilities is another question. Lebowitz expresses a common view when he says that, “Without this invariance the connection between phase space volume and probability would be impossible or at least very problematic” (1999b, 356).

had.<sup>52</sup> Likewise, the distribution need not have been uniform. There are other worlds, with similar dynamics, in which a non-uniform distribution is successful (worlds where nature favors compressed gases, say).<sup>53</sup> As far as we can tell, this is not our world. Consider that if we did observe a reliable preference for gases to stay clumped up in corners of their containers, then we would posit different probabilities, perhaps altering the dynamics and spacetime symmetries to reflect that preference. (As we might do if we reliably observed systems evolving toward certain spatial or temporal locations.)<sup>54</sup> We think that such preferences would show up in the dynamical behavior of ordinary systems. Since we don't observe these preferences, we reasonably infer that there are none—that we have posited the correct dynamics and probabilities.

There is a sense, though, in which this is the only measure that “respects the structure” of the dynamics. Guszczka (2000, Appendix II) proves that there is a unique measure on a symplectic manifold that is preserved by canonical transformations, the transformations preserving the Hamiltonian equations. (He proves that there is a unique (up to multiplicative constant) measure on a symplectic manifold that is finite on compact sets and preserved under all local canonical transformations.) And he shows that it is the measure associated with the standard volume

---

<sup>52</sup>Invariance, that is, is not an a priori constraint. Whether it holds depends on the theory; see Albert (2000, n. 8, 80). Branden Fitelson suggests this is the only distribution satisfying certain reasonable constraints, among them invariance. But I prefer not to impose such initial constraints, for it seems a different distribution could have been correct.

<sup>53</sup>Whether we would posit asymmetries in the spacetime, dynamics, or probabilities is a separate issue. The point is that we would infer some such asymmetries. Cf. Callender (2000).

<sup>54</sup>Compare Strevens' (1998) view that our probability inferences stem from observed symmetries plus knowledge of the laws, including which features the laws do, and which they do not, care about. The laws care about the shape of the die but not the number of dots painted on a side, for example; or so we (defeasibly) infer from the empirical evidence. Similarly, von Mises (1981, 73) says that, “no concrete case can be handled merely by means of an a priori knowledge of equally likely cases. It is always necessary to use more or less general results derived from observation and experience in order to determine which properties of the apparatus that we are using may influence the course of the experiments, and which properties are irrelevant from this point of view.”

element given by the symplectic form.<sup>55</sup> This gives a precise sense in which the standard measure is simplest relative to the statespace structure; it's the one preserved by the same transformations that preserve the dynamical laws. The invariant quantities of the laws dictate both the statespace structure and the natural measure on it. Again, it is a contingent feature of our world that the right measure to use is natural in this way.

These considerations are independent of your view on the nature of the probabilities. If you take them to be epistemic or merely instrumental, then the reason we use uniform probabilities for the purpose of statistical mechanics is that we generally choose the simplest formulation of a theory, all else equal. I happen to think that a realist and objective understanding of the probabilities is more natural; so that when we infer this distribution, we are inferring something about the world and its actual distribution of microstates.<sup>56</sup> I think that this yields a simpler, more unified theory, for reasons I expand on in section 5. Yet either way, this distribution is simplest, all things considered; and so, by our methodological rule, is the one we ought to infer, unless we were to get evidence otherwise.

Nor do these considerations depend on your view of laws of nature. The uniform distribution, if not a bona fide law, is an empirical generalization with many law-like features (it supports counterfactuals, figures in successful empirical explanations and predictions, and so on). If you have a best-system account of such generalizations, then posit this distribution on the grounds that it yields the simplest, most informative

---

<sup>55</sup>I thank David Malament for the reference. Following Guszczka (2000, Appendix II), for  $2n$ -dimensional phase space  $\Gamma$ , a canonical transformation is a diffeomorphism  $g : \Gamma \rightarrow \Gamma$  that preserves the symplectic form,  $\omega$ . A local canonical transformation is a diffeomorphism  $g : A \rightarrow B$ , for some neighborhoods  $A, B \subset \Gamma$  that preserves  $\omega$ . Guszczka proves that the natural measure associated with the volume element given by  $\omega$  on  $\Gamma$  is the unique measure finite on compact sets and preserved by the class of local canonical transformations of  $\Gamma$ . The result follows from Darboux's theorem and the fact that Lebesgue measure is the unique translation-invariant Borel measure on  $\mathbb{R}^{2n}$ . Darboux's theorem tells us that every pair of symplectic manifolds is locally isomorphic: within the neighborhood of every point, there are local (canonical) coordinates such that the symplectic form takes the canonical form. Two real symplectic manifolds (same dimension and signature) are locally identical: they can be mapped onto each other so that their symplectic structures correspond. See Arnol'd (1989, 230); Berndt (2001, 2.2); da Silva (2001, 8.1).

<sup>56</sup>Ismael (2009) argues that the probabilities must be so construed.

summary of the facts. If you are a non-reductionist about laws, then this is still the distribution to posit, on general methodological grounds. The non-reductionist will allow that we can be wrong in thinking these the right probabilities (even if we knew all the facts in the entire history of the world, say), yet maintain we are empirically justified in positing them.

You may accuse me of advocating a version of an indifference principle, something like: infer a distribution that is indifferent over the structure required by the dynamics. Call it an indifference principle if you like. To my mind, it is a far enough cry from the traditional idea that it is undeserving of the name. Unlike traditional indifference, the symmetry considerations here are not epistemic, a priori, or arbitrary; the (empirically confirmed) dynamics picks out the parameters over which to distribute probabilities uniformly. To the extent that there is any a priori element here, it is only to the extent that we take our general criteria of theory choice—simplicity, explanatory power, and the like—to be a priori to a degree; but only to that extent. Without taking a stand on the epistemic nature of those criteria, I claim that the current account is no more a priori than those. The justification is ultimately empirical, coming from the evidence we have for statistical mechanics and the particle dynamics, combined with the methodological principle that we do not infer more structure to the world and its physics than what the fundamental laws indicate there is.<sup>57</sup>

## 5. From statistical mechanics to coin tosses

Even if you follow me this far, you might wonder about the questions we started with. What about our usual inferences from symmetries to probabilities? What does statistical mechanics have to do with those?

I want to suggest that even though it may not seem like it—even though statistical mechanics seems to have nothing to do with die throws and coin tosses and the like—once we have the statistical mechanical distribution in place, it can explain the success of our everyday inferences from symmetries to probabilities. These inferences succeed because we

---

<sup>57</sup>Hence this does not rely on the type of Probability Principle dismissed by Davey (2008).

live in a world of which statistical mechanics is true.

Why think this is at all plausible? Let's look at a simple case of a coin toss. Before tossing a coin, we infer that each side is equally likely to come up. Repeated tosses confirm this prediction. Why does our initial inference succeed?

For simplicity, imagine that I am holding a coin balanced vertically on a table. The "toss" will consist in my letting go of the coin and its falling to the left (to land heads) or to the right (tails). In what follows, assume that the coin is governed by Hamiltonian dynamics.<sup>58</sup>

Think of the phase space of this coin. Consider the region corresponding to its initial macrostate. This region comprises the points representing microstates compatible with the coin's currently being in this location on the table, with its having this particular size and average temperature, with my hand's exerting this particular pressure on the coin, and any of the coin's other macroscopic features.

Out of all these microstates, think of the different possible combinations of positions and momenta for the particles that will result in a heads as opposed to a tails outcome when I let go of the coin. That is, think of all the different possible ways the coin's particles could be arranged—with different combinations of initial positions and momenta—so that the coin will land heads; and think of all the different possible ways the particles could be arranged so that the coin will land tails. Think in particular of the different possible combinations of momenta. In the idealized case we are setting up here, slight differences in the momentum of even a single particle will determine that the coin falls to the left as opposed to the right when I let go of it. (The classical dynamics is deterministic.<sup>59</sup>) Finally, place a uniform probability distribution over this phase space region—the region corresponding to the initial macrostate of the coin, the region containing the points representing microstates compatible

---

<sup>58</sup>I argued above that we should infer Hamiltonian mechanics. But notice that the right coordinates are going to be something like canonical coordinates. The classical equations are second order, requiring two sets of coordinates to characterize systems' states; in particular, one set of coordinates is the derivative of, or tangent to, the other. One set effectively characterizes the particles' relative locations, the other the "directions" or rates of change of their motions. That is what we will need here.

<sup>59</sup>Setting aside the cases of indeterminism: Earman (1986); Malament (2008); Norton (2008a).

with the coin's initial macrostate—just as statistical mechanics would tell us to do. (Ignoring, for now, the worry that the coin is not a type of system addressed by ordinary statistical mechanics. I return to this.)

You may be thinking that we must settle deep controversies in the foundations of statistical mechanics before we can continue. First, there is debate over whether the distribution is uniform over these microstates or a (proper) subset of them, viz. those that have evolved from lower entropy past states (even back to the initial state of the universe).<sup>60</sup> Second, there is debate over whether the initial distribution is taken over the phase space of the coin or the entire world; and you might think that the latter cannot be applied to the world's various sub-systems.<sup>61</sup> Third, you may worry more generally that an initial distribution won't tell us the probabilistic behavior of an ordinary system, whose state is confined to an energy hypersurface. Any lower-dimensional space gets zero measure on the standard volume measure over the whole of a system's phase space.<sup>62</sup>

We can leave these issues aside here. For a uniform distribution taken over the initial macrostate of the coin should yield the same probabilistic predictions as one that is conditionalized on the past, of the coin or world.<sup>63</sup> A probability distribution taken over the macrostate of the world at some time, combined with the deterministic dynamics, will induce a probability distribution over the world's possible microstates at any other time; conditionalize the initial distribution on the macrostate at the other time. Since any microstate of the world includes a microstate for any sub-system, this procedure will yield probabilities for the possible microstates

---

<sup>60</sup>As in the past hypothesis: Albert (2000). Earman (2006) argues against this proposal.

<sup>61</sup>Winsberg (2004a) argues that this requires a further posit, which we don't think is true; cf. Earman (2006, 420). A similar criticism is in Reichenbach's "branching systems" account (Sklar, 1993, 8.III); Winsberg (2004b) is an updated version of that idea.

<sup>62</sup>As in Frigg (2008b).

<sup>63</sup>That is, the same *future* predictions. It is in order to get correct inferences for the past that Albert (2000) argues for conditionalizing the usual distribution on initial low entropy. Another worry is that it is not true, in general, that the right distribution is uniform over a given macrostate: not all microstates relative to a set of macroscopic features are uniformly distributed. In reply, we can restrict the macroscopic features we conditionalize on (to the usual thermodynamic ones, say), or adopt Albert's view that once we include any relevant features, the uniform one (conditional on the past) will be correct.

of any sub-system at any time. Importantly, this will assign relatively *uniform* such probabilities. A standard assumption in statistical mechanics makes it plausible that, for an ordinary macroscopic system, a uniform distribution taken over the entirety of (a bounded region in) its phase space will be relatively uniform throughout just about any sub-region.<sup>64</sup> So that when we conditionalize the initial distribution on the sub-region (and renormalize), we get another distribution that is relatively uniform. Whatever your stance on these issues, we can safely assume a relatively uniform distribution over a system's initial macrostate.

(Intuitively, place a uniform distribution over a region in a higher dimensional space—say, a two dimensional plane—throughout which the “abnormal,” low-probability regions are scattered randomly in tiny clumps. Project this distribution onto just about any region in a (one dimensional) subspace—that is, any region that's not as small and scattered as the abnormal regions themselves—and renormalize. In other words, take a uniform distribution, on the standard measure, over the points that are in both the plane and the one dimensional subspace. Plausibly, this yields another uniform distribution, with similar probabilistic predictions for the future, such as extremely low probability for the abnormal microstates.<sup>65</sup> Either take a uniform distribution over the current macrostate region of a system, or over the microstates within this region that came from past lower entropy, and we should get a relatively uniform distribution.<sup>66</sup>)

---

<sup>64</sup>See, for example, Lebowitz (1993a,b,c, 1999b); Sklar (1993, 206); Pathria (1996, 20); Albert (2000). Thus, let  $M$  be a system's macrostate,  $\Gamma$  its phase space,  $\Gamma_M$  the region corresponding to  $M$ ,  $M_a$  the initial and  $M_b$  the later macrostate, and  $\Gamma_{Mab}$  the region of  $\Gamma_{M_b}$  that came from  $\Gamma_{M_a}$  (the microstates in  $\Gamma_{M_b}$  on trajectories coming via  $\Gamma_{M_a}$ ). Then “for systems with realistic interactions the domain  $\Gamma_{Mab}$  will be so convoluted that it will be ‘essentially dense’ in  $\Gamma_{M_b}$ ” (Lebowitz, 1993a, 10); “interactions the domain  $\Gamma_{Mab}$  will be so convoluted as to appear uniformly smeared out in  $\Gamma_{M_b}$ ” (1999b, S349). Indeed, this is part of what the arguments of Boltzmann and Gibbs, for grounding the second law of thermodynamics in statistical mechanics, plausibly show. For skepticism about its plausibility, see Frigg (2008b, 130-133).

<sup>65</sup>Indeed, it is because this yields similar probabilities that Albert, for one, argues that an initial universal distribution should ground our predictions of entropy increase to the future of individual systems, at the same time yielding correct inferences about their pasts. See the quotation from Callender in note 18; compare Figure 3.15 in Albert (2000). But see Frigg (2008b, 130-133) for skepticism about this scattering assumption.

<sup>66</sup>There is a standard way of restricting the Liouville measure to a subspace like the

Statistical mechanics says that, compatible with any given macrostate, including the initial macrostate of our coin, there are just as many microstates in which a given particle in the system is heading to the left as there are microstates in which the particle is heading to the right. There is a one-one mapping between microstates and their time reverses—the microstates with the time-reversed particle velocities—and for any microstate that realizes a given macrostate, so will its time reverse. A bit more precisely, according to statistical mechanics, the phase space volume of each set of such microstates, out of the total volume of the region representing the system’s initial macrostate, will be the same.

Keeping in mind that the momenta of the individual particles determine how the coin will land, the above means that, according to statistical mechanics, there are just as many ways for those initial momenta to be arranged, consistent with the initial macrostate, so that the coin will wind up tilting to the left when I let go of it, as there are ways for the initial momenta to be arranged so that the coin will tilt to the right. That is, half the phase space region corresponding to the coin’s initial macrostate is taken up by microstates such that, if the coin starts out in one of these, it will (deterministically) fall to the left; half the phase space region corresponding to the coin’s initial macrostate is taken up by microstates such that if the coin starts out in one of those, it will (deterministically) fall to the right. The uniform distribution over the phase space region representing the coin’s initial macrostate then says that any such “left-directed” microstate is *equally probable* as any such “right-directed” one.

In other words, out of our two possible outcomes—falling left or falling right; landing heads or landing tails—statistical mechanics says that each one is equally likely. Statistical mechanics counts the differences among possible *microstates* in such a way that they add up to an equal probability for each of the two possible *macroscopic* outcomes. Similarly, within a bunch of similar coin tosses, statistical mechanics says that their

---

energy hypersurface, with the volume of the surface the normalizing factor: Kac (1959, 63); Lebowitz (1993c); Pathria (1996, 56). You might wonder whether the simplicity considerations picking out the standard measure also favor the restricted one; the latter standardly depends on the Hamiltonian. Yet all we need is that the distribution is initially uniform, and that it yields an approximately uniform distribution over the energy hypersurface—a big, but relatively standard, assumption: note 18.

microstates will be distributed with about half the tosses starting out in “left-directed” microstates and half in “right-directed” ones. At the macroscopic level, this yields the prediction that approximately half the tosses will land heads and half will land tails, or that a given coin toss has a  $\frac{1}{2}$  probability of landing heads.<sup>67</sup>

Here, finally, is our payoff: that was our initial inference! We infer, on the basis of the macroscopic information we start out with, before observing any tosses, that the coin has a  $\frac{1}{2}$  probability of landing heads and a  $\frac{1}{2}$  probability of landing tails. We likewise 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. It suggests that our ordinary inferences from symmetries to probabilities succeed when the symmetries we observe match the symmetries in the statistical mechanical distribution. For when there is this correspondence between the macroscopic symmetries—between a heads and a tails outcome, say—and the symmetries in the distribution of fundamental microstates—as among the different combinations of particle momenta—a uniform distribution over the different possible fundamental states will yield a uniform distribution over the different possible macroscopic outcomes.

Figure 1 makes this intuitive.<sup>68</sup> Each point represents a different possible microstate for the coin, characterized by the initial height  $h$  of its center of mass above the table and the angle  $\theta$  between its surface and that of the table (where  $\theta = 0$  when the coin is parallel to the tabletop with heads up). The coin is released from rest.<sup>69</sup> The figure shows the different

---

<sup>67</sup>That is, in the limit of increasingly many tosses, there’s a high probability of heads about half the time; the set of microstates with this limiting frequency has measure one.

<sup>68</sup>Keller (1986) gives a similar argument for coin tosses without collision. See also the example of the Galton board in Maudlin (2007), although Maudlin uses it to argue that we do not need the full structure of a probability distribution (note 42).

<sup>69</sup>Suppose that immediately after release, the velocity of the coin’s center of mass is vertical and the angular velocity is parallel to the tabletop. These features will be preserved throughout the motion, since the forces on the coin (from gravity and the table) act vertically. Assume a uniform mass distribution in the coin, a smooth table, and negligible air resistance. While the coin is falling, its motion is free fall plus rotation at a constant angular velocity. The outcome, a function of the height of release, is then caused by the collision(s) of the coin with the table. See Kechen (1990a,b) for discussion.

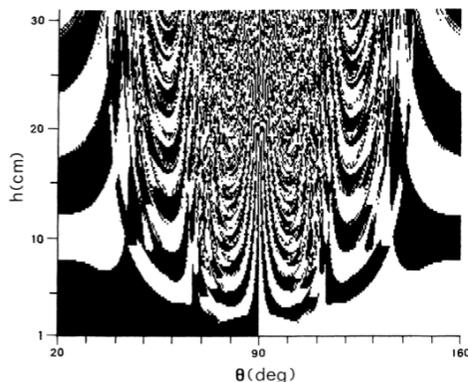


Figure 1: From Kechen (1990a,b). Dependence of final outcome on initial state of a coin released from rest above a table at height  $h$ .  $\theta$  is the angle between the coin and table surfaces;  $\theta = 0$  when the coin is parallel to the tabletop with heads up. Blackened regions represent states leading to heads.

possible states of the coin and shades the various regions according to their outcomes; regions shaded black represent microstates resulting in heads. Now place a uniform probability distribution (density), on the standard measure, over the points in the figure. This distribution will say that about half the initial microstates lead to heads (the black regions) and half to tails (white), or that the coin has a  $\frac{1}{2}$  probability of landing heads.<sup>70</sup> Indeed, imagine taking a uniform distribution over just about any reasonably-shaped subregion of the figure, and we will still get a probability of approximately  $\frac{1}{2}$  heads. In other words, plausibly, this distribution should assign probability  $\frac{1}{2}$  to a heads outcome even if we were to conditionalize on further features of the coin toss—that my hand can release it within a certain narrow range of velocities, for example.<sup>71</sup>

This is not to say that we must know what the statistical mechanical probabilities are. Our everyday inferences succeed even though we do

<sup>70</sup>Where above I argued that uniformly distributed momenta should yield a  $\frac{1}{2}$  probability of heads, here we have a similar argument from uniformly distributed positions.

<sup>71</sup>Compare Kechen: “Moreover, it may also be conjectured that  $H$  [the set of initial microstates leading to heads] is distributed uniformly within any region in phase space where the energy is large enough, and within any fixed region when coefficient  $e$  [of restitution] is sufficiently close to 1” (1990b, 1895).

not generally know about statistical mechanics. Rather, these inferences are successful because, and when, the observed symmetries align with the symmetries in the distribution of dynamical coordinates. The reason we make these inferences from symmetries to probabilities is our past experience, and this experience has been in a statistical mechanical world. We have learned from experience that systems' microstates are distributed in the way that the statistical mechanical distribution says they are.<sup>72</sup>

You might worry that something must be missing from the explanation of the coin toss. Imagine a machine with access to the exact initial microstate of the coin. Suppose the machine, on the basis of this information, can reliably toss the coin so that it always lands heads. Presumably, the machine can do this without any alteration to the dynamics.<sup>73</sup> This is a general point about statistical mechanical systems. A machine, or a Maxwellian demon, with knowledge of systems' microstates and the ability to act on this knowledge, could reliably cause systems to behave in anti-thermodynamic ways. Again, I take it that the job is to account for the use of these probabilities in explaining various macroscopic generalizations, to the extent that those generalizations hold. And we know that if we were Maxwellian demons, then the generalizations that we are trying to explain need not hold. It's not that we need an additional component in the explanation, in other words. It's that the explanation (and explanandum) hold only given that we don't have the epistemic and manipulative capacities of a Maxwellian demon. Otherwise, all bets are off.

The example of the coin is admittedly idealized. How the coin lands will depend on many other factors—the velocities of the surrounding air molecules, the angular velocities of the particles in my hand, and more besides. This should raise a skeptical eyebrow or two. Why think this procedure should yield the right probabilities once we include these real-life complications? The coin is not even a type of system ordinarily studied in statistical mechanics. Ordinary statistical mechanics talks about

---

<sup>72</sup>Compare von Mises (1981, 76): “most people, asked about the position of the centre of gravity of an unknown cube, will answer ‘It probably lies at the centre’. This answer is due, not to their lack of knowledge concerning this particular cube, but to their actual knowledge of a great number of other cubes, which were all more or less ‘true’.”

<sup>73</sup>I thank Jos Uffink for this question.

boxes of gas and cubes of ice, not dice and coins. Why think this should work for everyday systems like these?<sup>74</sup>

Eminently reasonable concerns. Yet I submit this as a plausible conjecture even so. As far as we can tell, there is no reason to think that statistical mechanics should not work for the particles in a coin, given its success for the particles in a gas. No difference in kind between the particles in these systems as far as the underlying dynamics is concerned; hence no difference in kind as far as the statistical mechanical predictions are concerned. And when we use the statistical mechanical probabilities to make predictions for something like a coin toss, we plausibly get the right results. It remains to show rigorously that statistical mechanics can be applied this generally, but it very well might. The empirical success of statistical mechanics suggests that the microstates of macroscopic systems are actually distributed in the way the standard distribution says—the microstates, that is, of *all* classical many-particle systems, boxes of gas and coins alike.<sup>75</sup> Given the success of statistical mechanics for these other

---

<sup>74</sup>Leeds (2003) is skeptical of any statistical mechanics that tries to carry the theory beyond the types of systems studied in ordinary research. A similar challenge could come from the view of Cartwright (1999) that we have no reason to infer that the laws of physics hold for ordinary systems, even granting their truth for systems we design in the laboratory. In reply, I think that the success (and truth) of the laws in systems we have studied indicates their general success (truth), until evidence shows otherwise. Another challenge comes from the view of Callender's (2008) that statistical mechanics is a special science; see North (forthcoming) for a reply.

<sup>75</sup>Here is one textbook: "Statistical mechanics is a formalism which aims at explaining the physical properties of matter *in bulk* on the basis of the dynamical behavior of its *microscopic* constituents. The scope of the formalism is almost as unlimited as the very range of the natural phenomena, for in principle it is applicable to matter in any state whatsoever. It has, in fact, been applied, with considerable success, to the study of matter in the solid state, liquid state or the gaseous state, matter composed of several phases and/or several components, matter under extreme conditions of density and temperature, matter in equilibrium with radiation (as, for example, in astrophysics), matter in the form of a biological specimen, etc." (Pathria, 1996, 1); original italics. Against such optimism, Callender cautions, "Outside of thermodynamics there is simply not a shred of evidence" that statistical mechanics underlies other, non-thermodynamic, macroscopic regularities (2008). Similar views are expressed in Leeds (2003). In reply, note that statistical mechanics has been applied widely, with continuing success. Where should we draw the line between thermodynamic and other kinds of macroscopic systems for which statistical mechanics will cease to work? At this point, we reach competing

types of macroscopic systems, assumed to be made up of the same kinds of particles, we have good reason to think the procedure should work even for a coin. Surprisingly, yet plausibly, the same distribution telling us that entropy increase is extremely likely in isolated thermodynamic systems also tells us the probabilities for the outcomes of coin flips. A bold hypothesis, to be sure, but one that is plausible all the same. A hypothesis that we should accept unless, or until, we find evidence to the contrary.<sup>76</sup>

Another view that may strike you as more plausible is that of Strevens (1998, 2003, 2005). Strevens also argues that underlying mechanical considerations help explain the success of our macroscopic probability inferences. He makes use of physical symmetries in the mechanism of a set-up and what he calls the ‘microconstancy’ of initial condition variables (roughly, that their values are macroperiodic). Yet his view gets by with the lesser constraint of a reasonably smooth initial distribution, not a uniform one. My proposal is more ambitious and correspondingly more prone to failure. Still, it may be correct; and if it is, it would yield a more unified overall account. According to it, we in fact have a wide variety of evidence for these probabilities—not only the different sorts of evidence we have for both statistical mechanics and the underlying dynamics, but also our everyday practice of die throws and coin tosses. The statistical behaviors of ordinary systems provide independent confirmation of the probabilities used in statistical mechanics. If successful, this account would explain more and in a unified way—including the success of Strevens’ own microconstancy condition. Since all things equal, we prefer theories that explain more on the basis of less, it seems worthwhile to treat it as a working hypothesis, unless, or until, we find evidence to the contrary.

You might worry that this requires understanding the statistical mechanical probabilities as objective, as corresponding to features in the world. I disagree. I do think that an objective construal yields a better

---

intuitions which I am unsure how to adjudicate. The present paper attempts to push the optimist line, but this needn’t convince the committed skeptic.

<sup>76</sup>Since the distribution over initial conditions will, when added to the deterministic dynamics, yield probabilistic predictions for every event in the history of a system, we can go further in saying that we ought to accept this until we find evidence to the contrary: such evidence would arguably amount to disconfirmation of statistical mechanics itself.

overall theory. For it lends itself to thinking of these probabilities as representing the actual statistical distribution of microstates; which we can then use to explain the success of indifference principles, when such principles succeed.<sup>77</sup> (In the same way that the scientific realist argues that realism best accounts for the success of our theories.) Yet the reasons for the uniform distribution remain whatever your metaphysics of its probabilities. At worst, take the above as suggesting why we use the uniform distribution, however we construe its probabilities. At best, take it as further argument that we treat these probabilities objectively, on the grounds that this view yields a deeper, more unified overall theory.

Finally, there is the question of how much the account will be able to explain. Does statistical mechanics underlie all successful uses of uniform probabilities? Do these dynamical considerations suffice to pick out the “right” probabilities for any of the familiar indifference principle cases (note 25)? I’m not sure. I do claim that this explains the use of uniform probabilities in statistical mechanics, and that this, in turn, explains the success of uniform probabilities in things like coin tosses. The extent of the account—can statistical mechanics explain *all* empirical statistical generalizations, as Albert (2000, 2008) argues? or only the predictions of ordinary statistical mechanical systems, as Leeds (2003) argues, albeit with the addition of systems like coins?—is for further investigation. Consider the current paper a modest attempt to nudge us closer to the “imperialist” view, as Callender (2008) calls it, of statistical mechanics as underlying many, if not all, of our probabilistic inferences about the world.

## 6. Conclusion

The statistical mechanical distribution at once explains the success of our everyday inferences from symmetries to probabilities, and justifies our symmetric probability assignments. We do not rely on an a priori principle to successfully infer the frequencies with which different possible outcomes occur. Initial symmetries relative to our epistemic state can’t tell us what the physical probabilities are. For they can’t even tell us what the fundamental possibilities are, let alone that any particular probability

---

<sup>77</sup>Such as the restricted use of the principle given by Castell (1998).

distribution over them will be successful. For similar reasons, neither must symmetries rationally constrain our initial credences.

Any seemingly a priori expectation we might have that the frequencies will match the symmetries in a situation really comes from past experience in a statistical mechanical world. We have learned from experience how systems' microstates are distributed, and we have updated our degrees of belief in what the initial probabilities are. We have also learned from experience what the dynamical laws governing systems' particles are, and from these what the relevant statespace structure is.

The symmetries in the world's dynamics and distribution of fundamental states, and the correspondence between those and the macroscopic symmetries we observe, explains the success of our everyday probabilistic inferences. Absent any such evidence, for truly prior credences, we can rely on symmetry in choosing an initial distribution. But so too can we choose some other, asymmetric distribution. Neither choice 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.<sup>78</sup>

## References

Abraham, R., & Marsden, J. E. (1980). *Foundations of Mechanics*. Reading, Mass.: Benjamin/Cummings, 2nd ed. First edition published in 1978.

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

Albert, D. Z. (2008). Physics and chance. Unpublished manuscript.

Arnol'd, V. I. (1989). *Mathematical Methods of Classical Mechanics*. New York: Springer-Verlag, 2nd ed. Translated by K. Vogtmann and A. Weinstein. First edition 1978.

---

<sup>78</sup>For discussion and comments, I am grateful to David Albert, Hartry Field, Branden Fitelson, David Malament, Ted Sider, Michael Strevens, Jos Uffink, students in a seminar at NYU, the Corridor group, and audiences at the University of Maryland, College Park and the Bellingham Summer Philosophy Conference and Philosophy of Science Association meeting in 2006, and two anonymous referees for this journal. I am especially grateful to Fitelson for comments on an earlier version of this paper at the Bellingham conference.

- Arntzenius, F. (2008). Credence and symmetry. Unpublished manuscript.
- Berndt, R. (2001). *An Introduction to Symplectic Geometry*. Providence, RI: American Mathematical Society.
- Bricmont, J. (1995). Science of chaos or chaos in science? *Physicalia*, 17, 159–208.
- Burke, W. L. (1985). *Applied Differential Geometry*. Cambridge, UK: Cambridge University Press.
- Callender, C. (1999). Reducing thermodynamics to statistical mechanics: The case of entropy. *Journal of Philosophy*, 96, 348–373.
- Callender, C. (2000). Is time ‘handed’ in a quantum world? *Proceedings of the Aristotelian Society*, (pp. 247–269).
- Callender, C. (2001). Taking thermodynamics (too) seriously. *Studies in History and Philosophy of Modern Physics*, 32, 539–553.
- Callender, C. (2004). Measures, explanation and the past: Should ‘special’ initial conditions be explained? *British Journal for the Philosophy of Science*, 55, 195–217.
- Callender, C. (2008). The past histories of molecules. Unpublished manuscript. Available at <http://philosophy.ucsd.edu/faculty/ccallender/ThePastHistoriesofMolecules.pdf>.
- Callender, C. (forthcoming). The past hypothesis meets gravity. In G. Ernst, & A. Hüttemann (Eds.) *Time, Chance and Reduction*. Cambridge: Cambridge University Press.
- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.
- Castell, P. (1998). A consistent restriction of the principle of indifference. *British Journal for the Philosophy of Science*, 49, 387–395.

- da Silva, A. C. (2001). *Lectures on Symplectic Geometry*. Berlin: Springer-Verlag.
- Davey, K. (2008). The justification of probability measures in statistical mechanics. *Philosophy of Science*, 75, 28–44.
- Dürr, D., Goldstein, S., & Zanghì, N. (1992a). Quantum chaos, classical randomness, and Bohmian mechanics. *Journal of Statistical Physics*, 68, 259–270.
- Dürr, D., Goldstein, S., & Zanghì, N. (1992b). Quantum equilibrium and the origin of absolute uncertainty. *Journal of Statistical Physics*, 67, 843–908.
- Earman, J. (1981). Combining statistical-thermodynamics and relativity theory: Methodological and foundations problems. In P. Asquith, & I. Hacking (Eds.) *Proceedings of the 1978 Biennial Meeting of the Philosophy of Science Association*, vol. 2, (pp. 157–185).
- Earman, J. (1986). *A Primer on Determinism*, vol. 32. University of Western Ontario Series in the Philosophy of Science: Reidel.
- Earman, J. (1989). *World Enough and Space-Time*. Cambridge, MA: MIT Press.
- Earman, J. (2006). The ‘past hypothesis’: Not even false. *Studies in History and Philosophy of Modern Physics*, 37, 399–430.
- Earman, J., & Rédei, M. (1996). Why ergodic theory does not explain the success of equilibrium statistical mechanics. *British Journal for the Philosophy of Science*, 47, 63–78.
- Ehrenfest, P., & Ehrenfest, T. (2002). *The Conceptual Foundations of the Statistical Approach in Mechanics*. New York: Dover. Translated by Michael J. Moravcsik. First edition in English published in 1959 by Cornell University Press. Originally published in 1912.
- Emch, G. (2007). Quantum statistical physics. In J. Butterfield, & J. Earman (Eds.) *Philosophy of Physics, Part B*, (pp. 1075–1182). North-Holland.

- Ercolessi, E., Morandi, G., & Marmo, G. (2002). Alternative Hamiltonian descriptions and statistical mechanics. *International Journal of Modern Physics A*, 17(26), 3779–3788.
- Feynman, R. P. (1998). *Statistical Mechanics: A Set of Lectures*. Colorado: Westview Press. Originally published 1972.
- Feynman, R. P., Leighton, R. B., & Sands, M. (2006). *The Feynman Lectures on Physics: The Definitive Edition*, vol. Vol. 1. Reading, Mass.: Addison-Wesley.
- Friedman, K. S. (1976). A partial vindication of ergodic theory. *Philosophy of Science*, 43, 151–162.
- Frigg, R. (2008a). Chance in Boltzmannian statistical mechanics. *Philosophy of Science (Proceedings)*, 75(5), 670–681.
- Frigg, R. (2008b). A field guide to recent work on the foundations of statistical mechanics. In D. Rickles (Ed.) *The Ashgate Companion to Contemporary Philosophy of Physics*, (pp. 99–196). London: Ashgate.
- Frigg, R. (forthcoming). Probability in Boltzmannian statistical mechanics. In G. Ernst, & A. Hüttemann (Eds.) *Time, Chance and Reduction: Philosophical Aspects of Statistical Mechanics*. Cambridge University Press.
- Gibbs, W. (1902). *Elementary Principles of Statistical Mechanics*. New Haven: Yale University Press.
- Goldstein, S. (2001). Boltzmann’s approach to statistical mechanics. In J. Bricmont (Ed.) *Chance in Physics: Foundations and Perspectives*, (pp. 39–54). Springer-Verlag. Available at [arXiv:cond-mat/0105242v1](https://arxiv.org/abs/cond-mat/0105242v1).
- Goldstein, S., & Lebowitz, J. L. (2004). On the (Boltzmann) entropy of non-equilibrium systems. *Physica D*, 193, 53–66.
- Guszcza, J. (2000). *Topics in the Foundations of Statistical Inference and Statistical Mechanics*. Ph.D. thesis, The University of Chicago.
- Guttman, Y. M. (1999). *The Concept of Probability in Statistical Physics*. New York: Cambridge University Press.

- Hagar, A. (2005). The foundations of statistical mechanics—questions and answers. *Philosophy of Science (Proceedings)*, 72, 468–478.
- Isham, C. J. (2003). *Modern Differential Geometry for Physicists*, vol. 61. Singapore: World Scientific Lecture Notes in Physics, 2nd ed.
- Ismael, J. (2009). Probability in deterministic physics. *Journal of Philosophy*, 106(2), 89–108.
- Jaynes, E. T. (1983). *Papers on Probability, Statistics, and Statistical Physics*. Dordrecht, Holland: D. Reidel.
- Kac, M. (1959). *Probability and Related Topics in Physical Science*. New York: Interscience.
- Kechen, Z. (1990a). Erratum: Uniform distribution of initial states: The physical basis of probability. *Physical Review A*, 42, 1012.
- Kechen, Z. (1990b). Uniform distribution of initial states: The physical basis of probability. *Physical Review A*, 41, 1893–1900.
- Keller, J. B. (1986). The probability of heads. *The American Mathematical Monthly*, 93, 191–197.
- Khinchin, A. I. (1949). *Mathematical Foundations of Statistical Mechanics*. New York: Dover. First English translation. Tr. by G. Gamow.
- Landau, L. D., & Lifshitz, E. M. (1980). *Statistical Physics*. Oxford: Butterworth-Heinemann, 3rd, part 1 ed.
- Lavis, D. A. (2005). Boltzman and Gibbs: an attempted reconciliation. *Studies in History and Philosophy of Modern Physics*, 36, 245–273.
- Lavis, D. A. (2008). Boltzmann, Gibbs, and the concept of equilibrium. *Philosophy of Science (Proceedings)*, 75(5), 682–696.
- Lebowitz, J. L. (1993a). Boltzmann’s entropy and time’s arrow. *Physics Today*, (pp. 32–38).
- Lebowitz, J. L. (1993b). Macroscopic laws, microscopic dynamics, time’s arrow and Boltzmann’s entropy. *Physica A*, 194, 1–27.

- Lebowitz, J. L. (1993c). Microscopic reversibility and macroscopic behavior: Physical explanations and mathematical derivations. *Physics Today*, 46, 1–20.
- Lebowitz, J. L. (1999a). Microscopic origins of irreversible macroscopic behavior. *Physica A*, 263, 516–527.
- Lebowitz, J. L. (1999b). Statistical mechanics: A selective review of two central issues. *Reviews of Modern Physics*, 71, 346–357.
- Leeds, S. (1989). Malament and Zabell on Gibbs phase averaging. *Philosophy of Science*, 56, 325–340.
- Leeds, S. (2003). Foundations of statistical mechanics—two approaches. *Philosophy of Science*, 70, 126–144.
- Loewer, B. (2001). Determinism and chance. *Studies in History and Philosophy of Modern Physics*, 32, 609–620.
- Loewer, B. (2004). David Lewis’s Humean theory of objective chance. *Philosophy of Science (Proceedings)*, 71, 1115–1125.
- Malament, D. (2008). Norton’s slippery slope. *Philosophy of Science (Proceedings)*, 75(5), 799–816.
- Malament, D., & Zabell, S. L. (1980). Why Gibbs phase averages work—the role of ergodic theory. *Philosophy of Science*, 47, 339–349.
- Marinoff, L. (1994). A resolution of Bertrand’s paradox. *Philosophy of Science*, 61, 1–24.
- Maudlin, T. (1995). Review of Lawrence Sklar’s *Physics and Chance* and *Philosophy of Physics*. *British Journal for the Philosophy of Science*, 46, 145–149.
- Maudlin, T. (2007). What could be objective about probabilities? *Studies in History and Philosophy of Modern Physics*, 38, 275–291.
- Milne, P. M. (1983). A note on scale invariance. *British Journal for the Philosophy of Science*, 34, 49–55.

- North, J. (2002). What is the problem about the time-asymmetry of thermodynamics?—a reply to Price. *British Journal for the Philosophy of Science*, 53, 121–136.
- North, J. (2009a). Structure in classical mechanics. Unpublished manuscript.
- North, J. (2009b). The ‘structure’ of physics: A case study. *Journal of Philosophy*, 106(2), 57–88.
- North, J. (forthcoming). Time in thermodynamics. In C. Callender (Ed.) *Oxford Handbook on Time*. Oxford: Oxford.
- Norton, J. (2008a). The dome: An unexpectedly simple failure of determinism. *Philosophy of Science (Proceedings)*, 75(5), 786–798.
- Norton, J. D. (2008b). Ignorance and indifference. *Philosophy of Science*, 75, 45–68.
- Pathria, R. K. (1996). *Statistical Mechanics*. Oxford: Butterworth-Heinemann, 2<sup>nd</sup> ed. First edition published in 1972.
- Penrose, O. (1979). Foundations of statistical mechanics. *Reports on Progress in Physics*, 42(12), 1937–2006.
- Penrose, O. (2005). *Foundations of Statistical Mechanics: A Deductive Treatment*. New York: Dover. First edition published in 1970.
- Pitowsky, I. (2006). On the definition of equilibrium. *Studies in History and Philosophy of Modern Physics*, 37, 431–438.
- Price, H. (2002a). Boltzmann’s time bomb. *British Journal for the Philosophy of Science*, 53, 83–119.
- Price, H. (2002b). Burbury’s last case: The mystery of the entropic arrow. In C. Callender (Ed.) *Time, Reality and Experience*, (pp. 19–56). Cambridge University Press.
- Prigogine, I. (1961). *Introduction to the Thermodynamics of Irreversible Processes*. New York: Interscience, 2<sup>nd</sup> ed.

- Schutz, B. F. (1980). *Geometrical Methods of Mathematical Physics*. Cambridge: Cambridge University Press.
- Shackel, N. (2007). Bertrand's paradox and the principle of indifference. *Philosophy of Science*, 74, 150–175.
- Sklar, L. (1973). Statistical explanation and ergodic theory. *Philosophy of Science*, 40, 194–212.
- Sklar, L. (1974). *Space, Time, and Spacetime*. Berkeley: University of California Press.
- Sklar, L. (1993). *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge: Cambridge University Press.
- Sklar, L. (2000). Interpreting theories: the case of statistical mechanics. *British Journal for the Philosophy of Science*, 51, 729–742.
- Sklar, L. (2001). Philosophy of statistical mechanics. *Stanford Encyclopedia of Philosophy*.
- Sklar, L. (2007). Why does the standard measure work in statistical mechanics? In V. F. Hendricks, K. F. Jørgensen, J. Lützen, & S. A. Pedersen (Eds.) *Interactions: Mathematics, Physics and Philosophy, 1860–1930 (Boston Studies in the Philosophy of Science)*, (pp. 307–320). Dordrecht, Holland: Springer.
- Strevens, M. (1998). Inferring probabilities from symmetries. *Noûs*, 32, 231–246.
- Strevens, M. (2003). *Bigger Than Chaos: Understanding Complexity through Probability*. Cambridge, Mass.: Harvard University Press.
- Strevens, M. (2005). How are the sciences of complex systems possible? *Philosophy of Science*, 72, 531–556.
- Tolman, R. C. (1979). *The Principles of Statistical Mechanics*. New York: Dover. Originally printed in 1938 by Oxford University Press.

- Uffink, J. (1996a). The constraint rule of the maximum entropy principle. *Studies in History and Philosophy of Modern Physics*, 27, 47–79.
- Uffink, J. (1996b). Nought but molecules in motion. *Studies in History and Philosophy of Modern Physics*, 27(3), 373–387.
- Uffink, J. (2004). Boltzmann's work in statistical physics. *Stanford Encyclopedia of Philosophy*. Available at <http://plato.stanford.edu/entries/statphys-Boltzmann/>.
- Uffink, J. (2007). Compendium of the foundations of classical statistical physics. In J. Butterfield, & J. Earman (Eds.) *Philosophy of Physics, Part B*, (pp. 923–1074). North-Holland.
- van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford: Clarendon.
- van Lith, J. (2001). Ergodic theory, interpretations of probability and the foundations of statistical mechanics. *Studies in History and Philosophy of Modern Physics*, 32, 581–594.
- Volchan, S. B. (2007). Probability as typicality. *Studies in History and Philosophy of Modern Physics*, 38, 801–814.
- von Mises, R. (1981). *Probability, Statistics and Truth*. Dover, 2<sup>nd</sup> ed.
- von Plato, J. (1994). *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*. New York: Cambridge University Press.
- Vranas, P. B. M. (1998). Epsilon-ergodicity and the success of equilibrium statistical mechanics. *Philosophy of Science*, 65, 688–708.
- Wallace, D. (2002). Implications of quantum theory in the foundations of statistical mechanics. Available at <http://philsci-archive.pitt.edu/>.
- White, R. (2005a). Epistemic permissiveness. *Philosophical Perspectives*, 19(1), 445–459.

White, R. (2005b). A new look at the old problem of induction. Unpublished manuscript.

Winsberg, E. (2004a). Can conditioning on the “past hypothesis” militate against the reversibility objections? *Philosophy of Science (Proceedings)*, 71, 489–504.

Winsberg, E. (2004b). Laws and statistical mechanics. *Philosophy of Science (Proceedings)*, 71(707-718).

Winsberg, E. (2008). Laws and chances in statistical mechanics. *Studies in History and Philosophy of Modern Physics*, 39, 872–888.