PROBABILITY IN DETERMINISTIC PHYSICS*

In a deterministic theory, one can deduce the state of the universe at any one time from the dynamical laws that govern its evolution and its state at any other time. In particular, one can deduce from the conditions that obtained at some early boundary condition for the universe its state at all subsequent times. It is a piece of well-received philosophical common sense that there is no room in a deterministic theory for objective probabilities. Arguments will be given below that this piece of philosophical common sense is mistaken, and that a probability measure over the space of physically possible trajectories is an indispensable component of any theory—deterministic or otherwise—that can be used as a basis for prediction or receive confirmation from the evidence.

Phase space is the central representational space in physics. It is the space within which all of the states pertaining to a system are represented, and it encodes all of the general information that we have about those states. To take an obvious example (less obvious examples will be given below) the fact that a pair of quantities are represented by different dimensions in phase space encodes the fact that they represent separate degrees of freedom. Physically possible histories are represented by trajectories through phase space and in a deterministic theory, when dealing with a closed system, there is (no more than) one trajectory through each point. Since the universe as a whole is the only truly closed system, this means there is one trajectory through each point in the phase space of the universe. Probabilities may be invoked in applications of theory, when the state of a system or the values of external variables are not known with full precision, but they are eliminable given complete knowledge. And they do not appear in the statement of the laws. The laws serve simply to identify the physically allowable trajectories. About Newtonian dynamics, Laplace famously said

> given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements

* Many thanks to Guido Bacciagaluppi, Huw Price, Elliott Sober, and especially Brien Harvey for comments. And I am grateful to audiences at Brown University, Sydney University, and the University of Western Ontario where earlier drafts of this paper were presented.

0022-362X/09/0602/89–108

> of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, and the past, would be present to its eyes.[1]

It is a short step from this to the idea that probabilities invoked in classical contexts are *subjective* rather than *objective*. If they represent anything, it is facts about our epistemic states, not facts about the world.[2] Karl Popper, and David Lewis, and other influential figures have endorsed this view, categorically rejecting any role for objective probabilities in deterministic contexts. In Popper's words:

> Today I can see why so many determinists, and even ex-determinists who believe in the deterministic character of classical physics, seriously believe in a subjectivist interpretation of probability: it is in a way the only reasonable possibility which they can accept: for objective physical probabilities are incompatible with determinism.[3]

And it is not just the philosophers; physics textbooks standardly assert that it is only with the advent of quantum mechanics where the laws take an explicitly probabilistic form that probability has to be taken seriously as part of the objective content of the theory.[4] Barry Loewer,[5] David Albert,[6] and others have been laboring to obtain recognition for the objective character of the probabilities given in the microcanonical probability distribution in Statistical Mechanics, and Elliott Sober has argued for the recognition of objective macroprobabilities.[7] But these have not gained a wide recognition for the role of probability quite generally. This paper will argue from general considerations that any theory capable of receiving confirmation from the evidence or generating expectation in a realistic setting incorporates a measure over state space as an indis-

[1] Pierre Laplace, *A Philosophical Essay on Probabilities* (1814), F.W. Truscott and F.L. Emory, trans. (New York: Kessinger, 2007), p. 4.

[2] Barry Loewer, expressing but not endorsing this reasoning, has written:
"If the laws are deterministic then the initial conditions of the universe together with the laws entail all facts—at least all facts expressible in the vocabulary of the theory. But if that is so there is no *further* fact for a probability statement to be about"—"Determinism and Chance," in *Studies in the History of Modern Physics*, XXXII, 4 (2001): 609–20, on p. 609.

[3] Karl Popper, *Quantum Theory and the Schism in Physics* (New York: Rowman and Littlefield, 1985).

[4] Quantum probabilities will be discussed in the sequel to this article. The laws of quantum mechanics specify not what *will* happen, given some initial segment of history, but what *might* happen with such and such probability.

[5] Loewer, "Determinism and Chance," and "David Lewis' Humean Theory of Objective Chance," *Philosophy of Science* LXXI, 5 (2004): 1115–25.

[6] Albert, *Time and Chance* (Cambridge: Harvard, 2000).

[7] Sober, "Evolutionary Theory and the Reality of Macroprobabilities," in Ellery Eells and James Fetzer, eds., *Probability in Science* (Chicago: Open Court, forthcoming).

pensable component of its objective content. Contrapositively, a theory devoid of such a measure is practically devoid of empirical content and practically incapable of receiving evidential support.

There is a very lively debate about the nature of probability (whether they are objective or subjective, whether they supervene on nonprobabilistic or nonmodal facts, and so on). A two-step approach will be favored here, separating what one says while working in the context of a physical theory where notions like law and probability and space and time are treated as primitive, from and what one says *about* laws or probability or space or time in a meta-scientific voice, where attempts at reduction or analysis can be offered. Aside from a few brief comments to address potential concerns of Bayesians, the discussion here will stop at the level of physical theory. The suggestion will be simply that we should recognize a probabilistic postulate as one of the standard components of a theoretical package, alongside the laws and space of states.

WHAT DETERMINISM ENTAILS

It was remarked above that there is only one physically possible trajectory through the phase space of a universe governed by deterministic laws. This means that if the state of the universe is known with perfect precision at any time, its state at any other can be predicted with certainty. In physics, however, when one is interested in the universe as a whole, one never knows its state with perfect precision, so it is represented not with a point, but with a finite volume of phase space. When one is not interested in the universe as a whole, but in an open subsystem of the universe, one is still dealing implicitly with finite volumes of universal phase space. For the state of an open subsystem of the universe is a partial specification of the state of the universe as a whole. Anything larger than a point-sized volume in the phase space of such a system corresponds to a finite volume in the universal phase space (extended in all dimensions external to it). When one is dealing with finite volumes of phase space, even in a universe governed by deterministic laws, one is no longer in a situation in which there is a single trajectory leading from one's starting point. There are multiple trajectories—typically, an infinite number—coming out of any finite volume of phase space. And dynamical laws that constrain only possibility—even deterministic laws that yield a *single* possibility for every point in phase space—are impotent to discriminate between possibilities. They will tell us which trajectories are possible, but will be silent about how to divide opinion among them.

In practice, when one is, for example, carrying out an experiment on some localized system in the lab, one can ignore the possibility that a comet will come at light speed and destroy the apparatus, that the air in the room will anti-thermodynamically collect in one corner, or that the pointer on our measuring device will be disturbed by a quantum tunneling event, even though there are models of the physical laws in which each of these occurs, because these possibilities have negligible or low probability. As possibility goes, they are on a par with those we take seriously (for example, that the result we observed tells us something about the system we are measuring). The epistemic symmetry is broken only by relative probability. It does not help to say there that many open subsystems are *approximately* closed, for probabilities are needed to make out *degrees* of closure. To say that a system is approximately closed is to say that the *probability* of outside influence is minimized.

If probabilities are needed to guide expectation under limitations in knowledge of the state of the universe, they are needed also to relate theory to evidence where it is a question not of *logical* compatibility, but *degree of support*. Logical compatibility with the data is a very weak constraint, certainly not strong enough to underwrite the theoretical choices that scientists actually make. In practice, theoretical choices between alternatives—from mundane, everyday inferences to choices between competing scientific theories—always involve assessments of likelihood. The Law of Likelihood says that given a choice between a pair of theories H_1 and H_2 compatible with a body E of evidence, one should choose the one that assigns the E a higher probability. Objective assessments of likelihood make use of a measure that the theory itself provides. In assigning a likelihood to H_1, one asks how special or contrived the initial conditions have to be by H_1's own lights to generate the regularities present in the phenomena. In assigning a likelihood to H_2, one asks how special or contrived or improbable the initial conditions have to be by H_2's lights to generate the regularities present in the phenomena. How much of what is observed in fact is inevitable, and how much is the result of coincidence or accident? If we observe a robust correlation between a pair of variables where a theory recognizes no connection, that lowers evidential support for the theory not because the theory rules it out as a logical matter, but because the theory assigns it a low probability. It is unlikely, but not impossible, that a pair of independent variables exhibit correlations. In the case of an indeterministic theory, we ask how likely are those regularities, given the chances described by the laws. If a theory assigns a high probability to one result, but whenever we carry out the experiment we always get the opposite, this disconfirms the theory not, again, because of a logical incompatibility, but because the theory assigns that result a low probability. The likelihood of the theory is a measure of how improbable the evidence is, by its own lights, and in making relative assessments of likelihood, we always employ a probability measure.[8]

Probability here is playing an indispensable mediating role between theory and evidence. Any theory that receives confirmation that favors it over empirically adequate alternatives on the basis of likelihood reasoning, or generates expectation that discriminates possibilities, has to tell us not just what is physically *possible* relative to a body of evidence, but the degree to which the various possibilities are to be *expected.* The kinds of regularities that are actually used, even in the most highly regimented experimental contexts (to say nothing of how things are out in the wild where expectations are formed about radically open systems on the basis of only the roughest information) are regularities about the probability with which a system that starts out in some finite, more or less well-defined, volume of phase space and is potentially subject to an indefinite number of potential extruding factors, ends up in another. Even if the intrinsic state of a local system is known with precision arbitrarily close to perfection, that state fills a finite volume of the phase space of the universe as a whole. In practice to derive expectations for a system *S* what is done is that one writes down everything that is known and thought to be relevant about *S*, about its environment, about relevant influences on it.[9] One comes up with a physical description as precise as possible in relevant respects.[10] That will carve out the volume of phase space of interest. He then takes his expectations for *S* from the behavior of systems whose states fall in that subspace, that is, systems that share everything that is known about *S*.[11] So, for example, if one wants to know how a particle in a certain state will evolve under a specified set of conditions, one identifies the subspace of the universal phase space that corresponds to particles of that kind, identifies the volume of that subspace

[8] Employing subjective probabilities in this capacity makes relative likelihood assessments a subjective matter and renders it all but void as an objective basis for choice between theories. If the measure is included in the scope of the theory as part of its content, the Law of Likelihood can provide a theory-independent measure of degree of fit with the evidence. See my "Likelihood Reasoning and Theory Choice" (manuscript).

[9] It is not clear that the notion of relevance can be made out without invoking probability, for *A* is a relevant factor if there is a non-negligible probability of influence from *A*. But put this aside.

[10] This is an application of the Principle of Total Evidence, usually credited to Rudolph Carnap, *Meaning and Necessity* (New York: Clarke, 1947).

[11] To know someone is a typical member of a flatly distributed population leaves one without information that favors one location over another, to know someone is a typical member of an unevenly distributed population is to know something that favors some locations, but not enough to locate them in a subpopulation. A typical American, for example, is more likely to live in New York than Tucson, but a typical molecule of water in the ocean has roughly equal probability in being located in any two volumes of equal size.

that corresponds to the relevant state, and applies the dynamical laws to determine how states that fall in that volume evolve under the relevant conditions. In doing so, one is effectively assigning to the system of interest the profile of the typical ensemble compatible with everything he knows about it. This fixes the values of known variables and randomizes over unknown variables. The probability of b on S calculated in this way is the probability of a random pick from a typical ensemble of S-type systems.

Even in a deterministic context, in which there is only a single trajectory through every point in phase space, unless one knows the state of the universe as a whole with perfect precision, probabilities are unavoidable. In practice one is warranted in ignoring most of the universe when considering the state of a local system because she has a probability distribution, at least implicit, that assigns a low probability to distorting influences from other parts of the world, that is, a probability distribution that defines what is meant by "random" and licenses the treatment of variables not represented in the model as random. That same probability distribution lets one interpret what she does observe. If one gets the same result every time an experiment is carried, although it is *possible* that this is pure coincidence, or that there is a string of experimental errors (the lab assistant misread the result, there is a freak event inside the measuring apparatus, and so on ...), these possibilities have low probability. The more reliably a result is observed, and the more robust it is with respect to variation in the surrounding circumstances, the more significant it becomes because, and only because, the probability of coincidence or accident is lowered if the result persists over time and across variation. The possibility of coincidence or accident remains, it is only their probability that is diminished. Again, the probabilities are playing a crucial mediating role between evidence and theory.

The probabilities here are independent of the dynamical laws. If one is looking at a finite volume v of state space, there will almost always be multiple trajectories coming out of v and leading in different directions. Some will go up and some will go down, and to form expectations, one needs to know *how many* of a typical finite set of systems whose states begin in v go this way and *how many* go that. If one wants to know the probability that a system whose initial state falls in v is in v^* after an interval I, that is to say, she calculates the probability that a random pick from a typical ensemble of systems whose states begin in v, ends in v^*. Dynamical laws that specify only what the physically *possible* trajectories are will not yield these probabilities. And so where there is more than one possibility, dynamical laws will not specify how to divide opinion among them.

Since, in the deterministic case, at least, the need for probabilities disappears when one possesses perfectly precise knowledge, it is easy to slip into thinking that the probabilities are subjective. If probabilities are only *needed* in forming expectation where there is ignorance, the probabilities must express facts about the epistemic states of belief-forming agents, rather facts about the systems they are forming opinions about. It is true that the probabilities reflect facts about the epistemic states of belief-forming agents, but there is a subtle equivocation in passing from the thought that probabilities provide a measure of degree of ignorance to the idea that they express nothing more than degree of belief. Credence is not the same as degree of ignorance. To characterize one's states as states that reflect a lack of information, one needs to invoke a notion of probability that is distinct from credence. A flat distribution over a partition of possibilities expresses ignorance only if the cells of that partition are equiprobable. Suppose I am attending a hosted dinner and I am uncertain about what will be served. If I assign equal credence to anything in the cookbook, that expresses ignorance about what my hosts will cook. But I could equally partition the possibilities for dinner into Sea Bass on quinoa pilaf on the one hand and anything other than Sea Bass on quinoa pilaf on the other, and an equal assignment of credence to the cells in this partition expresses quite specific information about the menu. Ignorance is indifference *between equally probable alternatives*, and the credences one adopts about S when one does not have specific information about S will inevitably invoke an objective, statistical notion of probability and express beliefs about which alternatives are equally likely to show up in a random sample. Credence that favors A over B or B over A, likewise, expresses the possession of information only if the expectation of A and B would be otherwise equal. A probabilistic postulate provides a base-line definition of statistical randomness that allows one to characterize credences as states of knowledge or ignorance.

These probabilities form part of the empirical content of the theory. It was remarked above that in forming expectations for a system S, one takes expectations from the behavior of typical systems compatible with everything known about S.[12] Facts about the behavior of typical ensembles of systems whose states span a finite volume of phase space are perfectly objective. They are probabilistically related to frequencies, and confirmable or disconfirmable by collecting statistics from ensembles. They are as objective, and as empirical, as facts about the

[12] This is the source of the link between single-case and in general probabilities, that is, between expectations for behavior of individual systems and frequencies of behaviors in ensembles.

phases of the moon, or the migration patterns of butterflies. "But surely," one might say, "deterministic dynamical laws dictate how any ensemble would evolve. If one knows the initial states of the systems of which it is composed, one applies the dynamical laws to determine their final states, and one is done." That is correct, the problem is not that one does not generally know how an ensemble of some specified composition will evolve, the problem is that one does not know *the composition of the typical ensemble*. This is a crucial difference. And notice that the same will be true in the indeterministic case. The dynamical laws there will generate a probability distribution over properties of interest for a system whose pure state is known, but they will not say how to mix probabilities for different states where there is uncertainty about which of them a system is in.

This is perhaps the deepest lesson of statistical mechanics. One does not get a probability distribution *out* of dynamical laws for free; one has to start with a distribution and evolve it forward. Most often an initial distribution is obtained by giving a flat distribution over the volume of phase space compatible with what we know about a system, tacitly applying a principle of indifference to states represented by equal volumes thereof. But we can reparameterize the space without affecting the topological relations among states (and hence without affecting Boolean relations among their extensions) or dynamical laws. And applying the same procedure under different parameterizations can give wildly inconsistent results. Talk about a flat distribution makes sense only once the metric of phase space has been fixed, and the practice of giving a flat distribution relative to a standard metric to reflect a lack of knowledge just expresses the assumption that states that occupy equal volumes under the standard metric are equally *intrinsically* probable, that is, that a random pick from a population that spans any two subspaces of equal volume is equally likely to yield a system in one as in the other.

The role that the measure over phase space was playing in generating probabilities in statistical mechanics was exposed when it was demonstrated that one can hold fixed the laws governing the evolution of individual microstates and derive quite different probabilities for evolution on the macrolevel (where macrolevel descriptions are just less than perfectly precise descriptions of microbehavior) by inflating some macrostates and deflating others. Statistical mechanics taught us to see the metric as an independent theoretical variable—one of the things that can be played with in bringing theory into line with the evidence. What one is playing with when one entertains different hypotheses about the relative volumes different states take up in phase space is hypotheses about the composition of typical ensembles. The naïve expectation would be that the typical ensemble is a roughly even composition of macrostates.

The hypothesis of microuniformity, by contrast, gives us a composition so heavily weighted toward equilibrium states that almost all systems in almost all possible universes are in equilibrium.[13] Under application of the dynamical laws, the hypothesis of microuniformity yields a migration towards higher entropy states from lower entropy states as byproduct of the difference in phase space volume of states of high and low entropy. Overall, there is almost no migration since virtually all systems are in states of high entropy. But if our universe is (as the Past Hypothesis invites us to suppose) one of those very rare universes in which there are a great many systems in low entropy states, migration will be observed, and the migration observed will virtually all be in the same direction.

So to reconnect with the themes above, according statistical mechanics the expectation that one should see as many entropy decreasing processes as entropy increasing ones depends on a faulty assumption: namely, that all macrostates have equal statistical probability in the sense that a random pick from a typical ensemble is equally likely to yield a system in any chosen macrostate. There is dispute about the status and interpretation of statistical mechanical probabilities. But one thing that can be uncontroversially learned from that example is the purely formal fact that one can get radically different expectations by adopting different metrical assumptions, even while holding the dynamical laws fixed. Dynamical laws alone, even deterministic laws that dictate the evolution of a closed system given precise information about its state, do not constrain the division of opinion when initial information is incomplete or less than perfectly precise, or when dealing with an open system with limited knowledge of extruding factors. Talk of dividing opinion evenly among the possibilities—which here takes the form of spreading it uniformly over the part of phase space whose boundaries reflect our information about the system—is well defined only once the metric has been fixed, and only because the metric is functioning tacitly as a measure of the relative probability that a random pick from a population whose states fall within two specified subspaces will yield a system in either.

Although the lesson was one that historically, emerged from statistical mechanics, it has nothing specifically to do with dynamics, and

[13] At all times, the expectation is high entropy. If one is in one of those rare universes in which there are many systems in low entropy states, they are almost all evolving to states of higher entropy. Higher entropy states are intrinsically so much more likely that if one is in a high entropy state one is almost certain to stay there, and if one finds oneself in a low entropy state, one is almost certain to evolve into a higher one. The composition of any random ensemble of systems is heavily weighted towards high entropy states.

the point can be illustrated with simpler examples. The need for an independent metrical assumption has to do with mixing, and can be made in nondynamical terms. Whether dealing with points in phase space, cells in a partition, or categories in a taxonomy, dynamical laws that give profiles for individual cells need to be supplemented with a comparative measure of the sizes of the populations that occupy cells, something that specifies how systems distribute themselves among cells and can be used to yield the probability that a random pick (from the population as a whole, or a subpopulation drawn from a specified collection of cells) will yield a member of the cell in question. One gets different expectations depending on which is regarded as partitioning them into equally probable classes.[14] It is much clearer in finite examples where one can simply count up the inhabitants. In these cases, the size of a cell is given by the number of its inhabitants and probability corresponds to proportion. Let us take the set of people in the world and partition them by *country of birth*. Suppose that nationality is deterministic with respect to language and car make, so that the country in which a person is born determines the language she speaks and the kind of car she drives. Everybody born in the United States speaks American and drives a Ford, everybody born in France speaks French and drives a Peugeot, everyone born in Samoa speaks Samoan and walks to work, and so on. In forming expectations about a fellow of whose country of birth one has no specific knowledge,[15] if one naively divided opinion evenly among the possibilities for country of birth, she would expect him to be equally likely to speak Chinese as Samoan, equally likely to drive a Ford as a Peugeot, and so on. She would be mistaken, and mistaken for the simple reason that these probabilities would not reflect statistics of the properties of interest in the world population. Taking the number of people in each cell as a measure of its size, let us say that a partition is *even* with respect to a given population just in case it divides the population into cells of equal size. The problem with the partition by country is that it is not even with respect to world population. A typical randomly chosen ensemble of world

[14] It does not matter whether we are talking about dynamical laws or a function from cell number to properties of interest. It does not matter whether we are talking about the deterministic or indeterministic case. And it does not matter whether we are talking about finite or infinite populations. In the finite case, size means number; in the infinite case, it means measure, but comparisons of size are no less dispensable in the second case than in the first.

[15] That will, in its turn, generate profiles for people of whom we have information represented by a probability distribution over some finite region. The typical global individual provides a kind of standard or reference point, departures from typicality represent knowledge.

individuals is not composed of nationalities in equal proportion, and probabilities guided by assigning equal probabilities to nationalities will not reflect statistics in the global population. Consider by contrast the partition of the same population by date of birth. This is not perfectly even with respect to population, but it is more nearly so.[16] On any given day, the number of people born in China will radically outweigh the number born in Samoa, the number of people born in the United States will radically outweigh the number born in France. The expectations formed by spreading opinion evenly over the cells of this partition will be quite different than those formed by spreading it evenly over the cells of the partition by country, and will be more in line with the actual frequencies.

So, to repeat, when forming expectations for a system S about which one has imprecise knowledge, one writes down everything known about S and takes expectations for S from typical ensembles of systems that share those properties. A complete profile for members of each cell does not give a complete profile for populations drawn from multiple cells until the proportion that individual cells contribute to the combined population is specified. One can derive different expectations for the global population, or any subpopulation that spans cells, by fiddling with how many elements each cell contributes to the larger population. Talk of metrics for phase space is just a generalization of this. The probability with which an arbitrary pick from the population yields a system in a particular state is proportional to the volume it occupies in phase space. One can use whatever numbers one likes to parameterize phase space, but nonmetric preserving reparameterizations will alter expectations under this procedure for forming them. What is the right parameterization? That will depend on how systems actually distribute themselves over phase space. One has to make some metrical assumption before expectations can be formed. Expectations derived from a flat distribution over a space whose metric does not reflect the way that the systems actually distribute themselves leaves one with probabilities distorted in favor of sparsely populated cells.

Another example of failing to appreciate the need to take account of the relative sizes of the cells of a partition in forming expectations, interesting because it actually reverses conditional expectations for a combined population that hold individually for each of the components. The University of California, Berkeley, was very famously sued for bias against women applying to grad school. The admission figures showed that men applying were more likely than women to be admitted,

[16] Assume, here, that births distribute themselves roughly evenly over the year in each country.

and the difference was significant. But examining individual departments found most departments actually showed a slight bias against *men*. There is no paradox here. The explanation was simply that women tended to apply to departments with low rates of admission, while men tended to apply to departments with high rates of admission. The population as a whole (people who apply to grad school) was composed of a set of subpopulations (people who apply to the English department, people who apply to the Physics department, and so on). In each subpopulation, the rate of acceptance for men was lower than that for women, but in the combined population, the rate of acceptance for women was significantly lower than that for men because the composition of the ensemble of women applicants was heavily weighted towards departments with low rates of acceptance, while the composition of the ensemble of male applicants was weighted towards departments with higher rates of acceptance, with the result that men had a higher overall probability of being accepted. It is as though they were in the same casino, but the women tended to play games with a lower chance of success than the men. You can always increase your intrinsic probability of winning by playing the game with better chances. The correlation between being a woman and being rejected was an artifact of the composition of the ensemble of female applicants, in just the way that the higher relative probability of speaking English over Samoan is an artifact of the composition of the ensemble of persons, and in just the way that the higher relative probability of entropy increasing processes over entropy decreasing ones is an artifact of the composition of typical ensembles in worlds that start out in a state of low entropy.

Another example of the way that an uneven partition can artificially favor some elements is that I am the most hyperlinked 'Jenann' on the internet. If one googles 'Jenann', links to me, or pages of mine, come up first, second, third, and fourth. But David Chalmers is not the most hyperlinked 'David' on the internet, and Saul Kripke is not the most hyperlinked 'Saul' on the internet. That does not mean that I am more famous than David Chalmers or Saul Kripke. The reason is that there are fewer Jenanns than there are Davids or Sauls (at least among the relevant population), and so I am more intrinsically likely to be most linked among my conomials. I am a big fish only relative to the size of my pond. The pond of my conomials is ever so much smaller than the pond of Davids or Sauls.

PRINCIPLE OF INDIFFERENCE

Some will have already spotted a connection with that notorious principle of probabilistic reasoning that has fallen into ill repute: the

Principle of Indifference (PI).[17] PI enjoins that given a set of possibilities and no specific reason for thinking one obtains rather than another, they should be assigned equal probability. It captures the intuition that, for example, if one has a six-headed die, one should assign probability 1/6 to each face on any given roll, or if one is presented with a spinning wheel labeled with numbers in the interval $0>x>1$, one should assign equal probabilities to obtaining a number greater than 0.5 and less than 0.5 on a random spin of the wheel. There is a wealth of counterexamples to PI that derive inconsistent results by applying the principle to the same events under different descriptions or considered as part of different partitions. Here is an especially simple one:[18] A factory produces cubes with side lengths up to 2 cm. What is the probability that a randomly chosen cube has side length under 1 cm? Apply PI to side length, one gets the answer 1/2. A factory produces cubes with face area up to 4 cm^2. What is the probability that a randomly chosen cube has face area under 1 cm^2? Apply PI to face area and one gets the answer 1/4. These are the same problem under different descriptions. One gets contradictory answers because application of PI to side length is tantamount to assuming that the factory produces cubes whose sides show a uniform distribution in length, whereas application of PI to face area is tantamount to assuming that the factory produces cubes whose faces show a uniform distribution in area. But a production process uniformly distributed in side lengths produces a different distribution of cube volumes than a process uniformly distributed in face areas.

The spread of cube volumes depends on whether one assumes uniformity over side length or face area, and which of these assumptions is correct is a question of fact. It is a matter of which partition is in fact even, determined by statistics for the cubes the factory actually turns out. PI will give the right results if *and only if* applied to the partition that actually divides the cubes coming out of the factory into

[17] This is also relevant to the Doomsday argument and Shooting Room set-up. In the Doomsday case, the natural metric imposed by supposing a person to be equally likely to live at any given time, and in the shooting room the metric imposed by regarding an arbitrary player to be equally likely to be playing in any given cycle of the game are what generate the mistaken expectations. Partitioning the population by date of birth, or partitioning players by game cycle, does not divide it into equally sized cells and expectations derived by projecting relative frequency in population overall is weighted in favor of later cells. An arbitrary member of the population is much more likely to live to see Doomsday and an arbitrary player of the shooting game is 9 times more likely to win than to die, though the chance of Doomsday at any one time is as likely as at any other, and the chance of losing in any cycle of the shooting game is just as likely as that of winning.

[18] Derived from Bas van Frasssen, *Laws and Symmetry* (New York: Oxford, 1989).

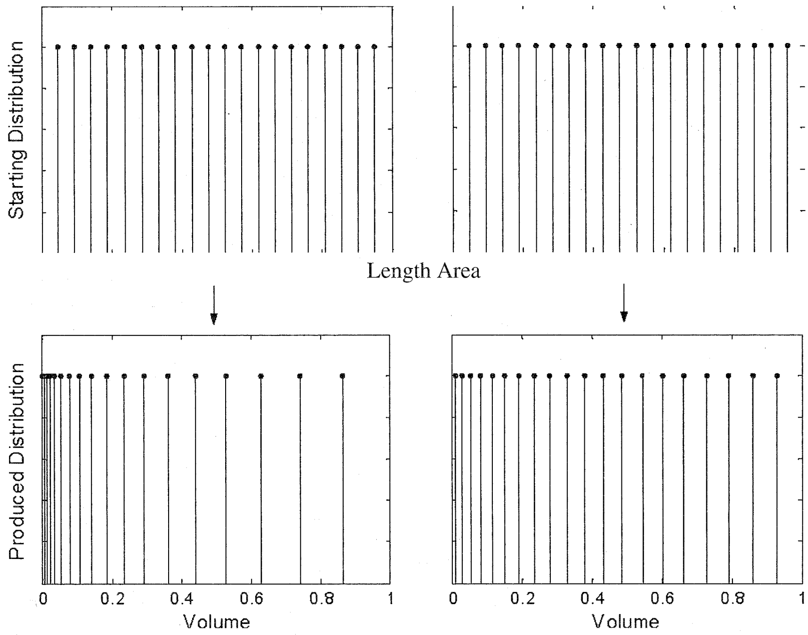

Figure 1: The top two panels represent uniformity in length and area. The bottom two panels show how these uniform distributions change when mapped to volume. A process uniform in length produces a different set of volumes than a process uniform in area.

equally sized classes. That is an unavoidably empirical question. There is no more a way of identifying the even partition from logical principles than there is of using such principles to divine what the factory heads decide at this year's corporate convention. Given a random spin of the wheel whose edges are labeled between 0 and 1, should one spread probability uniformly over equal intervals? Or should one use some other function? There are as many ways of spreading probability over the interval $0<x<1$ as there are real-valued functions whose arguments in that interval sum to 1. Which is correct depends on whether the wheel spins at uniform speed, and how numbers are actually distributed around the wheel. And, again, that is an unavoidably empirical matter. What is right for one wheel will be wrong for another. When PI is applied to reflect a lack of specific information about a system, it can be done in a well-defined manner only against the background of general information about how properties of interest distribute themselves across the population from which the system is drawn. Consider a circle of finite volume on which all colors in the visible spectrum are represented. If asked for the probability that a randomly chosen surface will have a color that falls on the top half of the wheel, many will automatically answer ½. This response is correct assuming the standard representation of the spectrum on which colors separated by equal wavelengths are separated by equal volumes on the circle. But it is perfectly possible to crowd all colors except for a particular shade of deep purple into the bottom half of the wheel. Nothing said when the question was introduced specified how colors were to be distributed around the circle, and the answer would be wrong on this representation. The ways families of properties are represented visually and imaginatively almost always incorporate assumptions, rarely explicit, about statistical probability, that is, the probability that a random pick from a typical ensemble of systems of the relevant type will have the property or properties of interest. It is these assumptions that are made explicit when their implications are drawn out in applications of PI. Richard Von Mises, discussing Bertand's chord paradox, says the same thing about the absence of formal criteria for identifying the right metric. By 'coordinate system', here, he means parameterization:

> ... the assumption of a 'uniform distribution' means something different in different co-ordinate systems. No general prescription for selecting 'correct' co-ordinates can be given, and there can therefore be no general preference for one of the many possible uniform distributions.[19]

Although one is rarely conscious that one is doing it, one chooses parameters that reflect one's metrical assumptions, and it is really the metrical assumptions built into the choice of parameters that underwrites the practice of assigning uniform distribution to reflect ignorance. The metrical assumptions are wholly empirical, and the practice is only successful to the degree that they are correct. Assigning probabilities to events to reflect a lack of specific information, that is, a lack of information that singles the system of interest out from the crowd, requires some general knowledge about the composition of the crowd. In physical contexts, this kind of general knowledge is embodied in the metrical structure of phase space. Specific information gets used to carve out a volume of phase space, but when specific information gives out, one gives a flat distribution over the remaining volume and general knowledge kicks in. If metrical structure faithfully reflects statistical probabilities, when it is interpreted by the principle that a random pick is equally likely to yield a system in any two equal volumes of the space, indifference over the volume of phase space whose boundaries

[19] Von Mises, *Probability, Statistics and Truth* (New York: Macmillan, 1957), pp. 80–81.

reflect the extent of our specific knowledge will give us probabilities that reflect statistics in the relevant population.[20] We have to make *some* metrical assumption before we can form expectations, expectations derived from a flat distribution over an uneven partition or a space equal volumes of which do not represent statistically equiprobable states is going to leave us with distorted probabilities. In a finite example, there is an easy way of identifying the even partitions; we count the occupants of cells. In an infinite population, where proportions are replaced by measures and there is no intrinsic measure of size of classes, identifying the even partition requires all of the tools of modern statistics, and there are no logical guarantees that the statistics in any finite sample will be representative.

HIGHER ORDER IGNORANCE

Of course one can be ignorant of general matters too, and general ignorance will compound with lack of specific information when probabilities are calculated. General ignorance is ignorance of how properties of interest distribute themselves across the population whose boundaries reflect the extent of specific knowledge. It is represented by mixing probabilities drawn from different hypotheses about this distribution.[21] So, in the example above, if it is not known how the factory makes cubes, or it is not known how numbers are distributed around the spinning wheel, but one knows how to partition the alternatives into equiprobable classes, PI can be applied at a higher level by forming a mixture weighted to reflect higher order indifference between equiprobable alternatives. But one does not always know how to form an even partition of alternatives, and in some cases it does not make sense. This kind of higher order ignorance is complex. It remains, however, that the first order content of a physical theory, insofar as the theory is used to generate expectation in the face of ignorance, includes a probabilistic postulate which serves as part of the background against which the impact of incrementally obtained

[20] We have to be careful to understand that the indifference is not over possible outcomes for an experiment, but indifference over outcomes generated by the same method. Incorporate everything you know about the system into the process that generates the outcome, and think of selection as a random choice from outcomes generated by the same process. So, for example, if you are assigning probabilities to a dart thrown by John hitting the center circle, you take all of the specific knowledge you have of the process, and let the probabilities reflect the distribution of outcomes in that class (for example, darts thrown by John from a certain distance in bad light after he has had three beers, while he is talking to buddies, and so on).

[21] 'Mixing' is a technical notion, the mixture of probabilities P and P^* drawn from different sample spaces is $(\alpha P + \beta P^*)$ where α and β are the relative probabilities of the sample spaces from which they are drawn.

information is assessed. Phase spaces and state spaces are what statisticians refer to as 'sample spaces'. They embody a great deal of general knowledge of an empirical nature, and that knowledge is relied on when a theory is used to form expectation on the basis of limited specific information.[22]

THE LAPLACIAN ARGUMENT

If what preceded is correct, there ought to be a gap in the knowledge of the Laplacian intelligence. There is. It is revealed by the fact that the Laplacian intelligence cannot answer problems of the form: 'What is the statistical probability with which a system in state A at t ends up in state B at t^*, given C?' where A and C together fall short of a complete specification of the microstate of the universe, and B is fine-grained enough not to permit a 1 or 0 value. By hypothesis, the Laplacian intelligence knows the microstate of the universe, and he knows the dynamical laws, and whenever he tries to make predictions about the behavior of a thermodynamic system, he evolves forward the microstate, so he never has to worry about probabilities and his predictions are never wrong. His extraordinary specific knowledge keeps him from confronting problems of this form. Problems of this form, however, are the rule for us, and the question is not whether the Laplacian intelligence confronts them, but whether he has what he needs to provide answers. The thermodynamic example demonstrated that he cannot squeeze answers out of the dynamical laws alone, for one can hold the laws fixed and generate radically different answers to questions of this form by adopting different probabilistic postulates. Looking at frequencies will provide *evidence* for probabilities, but as a logical matter they do not determine the probabilistic facts. Perhaps, as those that defend Humean reductions suppose, facts about probabilities are included implicitly in the manifold of categorical fact, not directly in the frequencies but along with the laws as part of the best overall systematization of the manifold.[23] That may be correct, but it is separate from the thesis that has been defended here: namely, that any systematization that can be used as a basis for prediction or receive confirmation from the evidence will incorporate a measure over phase space, and

[22] The boundary between general and specific knowledge is conventional. What counts as specific knowledge about a particle in a context in which we are using a state space for generic particles (for example, that it is negatively charged) is trivial (has probability 1, for every particle) in a context in which we are using a state space for electrons.

[23] Loewer, "Determinism and Chance," *op. cit.*, and Lewis, "Humean Supervenience Debugged," *Mind*, CIII, 412 (1994): 473–90.

the measure will be just as real and objective as, but not reducible to, the physical laws.[24]

BAYESIANISM, AND THE BEST SYSTEMS ANALYSIS

A couple of residual remarks are warranted to draw connections to other work in the literature. Loewer's discussion of statistical mechanical probabilities in "Determinism and Chance" is in some ways in accord with the position that has been defended here. His argument begins with the so-called Best Systems Analysis of laws. According to the Best Systems Analysis, what makes a certain claim about the world a law is that it is part of a theoretical package that overall gets the best fit with the evidence. The idea is that the correct physical theory will be that systematization that achieves the best overall fit with the global pattern of local matters of particular fact. He argues on this basis that the microcanonical probability distribution associated with statistical mechanics has a claim to be included in a theoretical package alongside the laws because when it is added to the laws, it vastly increases the informativeness and fit of the best system. He writes

> If adding a probability distribution over initial conditions—in our examples over initial conditions of the universe—to other laws greatly increases informativeness (in the form of fit) with little sacrifice in simplicity then the system of those laws together with the probability distribution may well be the best systematization of the occurrent facts.[25]

I am in full accord that we should regard a probability postulate alongside the laws as part of the theoretical package. The differences between Loewer's position and mine are these: (i) I have argued that we should do this quite generally, not just in the context of statistical mechanics, (ii) I think we should take $\Pr_G(A/B)$ as the basic object, rather than a distribution over initial conditions. $\Pr_G(A/B)$, recall, is defined as the probability that a random pick from *B*-systems would yield an *A*. It can be generated from a distribution over initial conditions in the context of deterministic laws, but does not require the idea of an initial state and has applicability also in indeterministic contexts, (iii) I am resistant to regarding $\Pr_G(A/B)$ as a *law* for reasons that are independent of what was said here, and finally (iv) where Loewer thinks that including a probabilistic postulate increases fit

[24] What our physical theories do is "solve" the data for generalizable patterns. I am arguing here that those generalizable patterns are not just the regularities embodied in physical laws, but patterns of migration over state space, not derivable from the laws without invoking a measure over state space, patterns that tell us how a group of systems, distributed across a finite volume of state space redistribute themselves over time.

[25] Loewer, "Determinism and Chance," *op. cit.*, p. 10.

with the evidence, I think that without a probabilistic postulate, we do not have an objective measure of fit with the evidence that goes beyond logical compatibility.[26]

Note that Bayesians can quite comfortably accept the thesis that a probabilistic postulate should be recognized as a standard component of a physical theory, even in deterministic contexts. One of the defining features of Bayesianism is the thesis that all probability is ultimately subjective, but there is some equivocation about what this actually entails. Subjective probability measures—in particular, the conditional probabilities that govern how belief is updated incrementally in response to evidence—encode a lot of information about the probabilistic relationships among events. A Bayesian can view a physical theory as nothing more than a candidate credence function, that is, a regimented rendering of the conditional connections among belief that govern the updating of opinion. This is what is known in the literature as an *expert function.* Insofar as one is viewing a physical in this way, she will have a special reason to recognize a probabilistic postulate as part of its content, for without a probabilistic postulate, a theory does not constitute a well-defined credence function.

CONCLUSION

Arguments have been given here for recognition of a type of probability that is more basic than chance, that is not intrinsically dynamical, and that is a practically indispensable component of deterministic and indeterministic theories alike: $\Pr_G(A/B) =_{\text{def}}$(the probability that a random pick from B-systems will generate a system in A). When A and B are dynamical states, $\Pr_G(A/B)$ gives the transition probability from B to A. It also gives the transition probability from A to B, and the nondynamical probability that a system has one property, given that it has another, for example, the probability of being green, given the property of being an emerald or the property of being in microstate A, given the property of being in macrostate B. It encodes virtually all of the inductively important information in a physical theory, everything that cannot be strictly deduced from known fact *via* the laws. If one knows enough that the physical laws leave only one possibility, the measure is not needed. But absent perfect and precise knowledge, there is always an indefinite number of possibilities and the laws do not discriminate among them. They do not say which of a set of possibilities is *more or most* possible, and so they do not tell one how to divide opinion between them. As for the suggestion to

[26] Any use of subjective probabilities will mean that agents with different priors will differ in their assessment of relative fit with the evidence.

divide opinion equally; 'dividing it equally' is not well defined until a metric has been specified. What is really meant when one is instructed to divide opinion equally is 'distribute it evenly *among equally probable alternatives*'. That just emphasizes the ineliminability of Pr_G. If one is looking to predict where one does not have complete information, or to intervene where one does not have precise control, one needs probabilities, and the probabilities have to be objective in the sense that they have to reflect the statistics in the population of interest. The more detailed one's specific information about the system, and the better her knowledge of the statistics, the more likely her success.

In arguing that a probabilistic postulate should be made one of the standardly recognized components of a theoretical package, it is reasonable to ask whether this is intended as a redescription of assumptions implicit in practice or a recommendation to begin including a probabilistic postulate in statements of theory. The answer is that there is both an observation and a recommendation. The observation is that insofar as one derives expectations from theory in a real setting, one is in fact invoking probabilities. The recommendation is to bring the often tacit, intuitive measures employed in practice into the open by including them in the explicit content of a theory, where their role in generating expectations is made transparent, and where they can be subjected to scrutiny and compared with alternatives.

J.T. ISMAEL

Centre for Time, University of Sydney
and University of Arizona